Cosmology and the Local Group in ΛCDM and Modified Gravity

Michael Alexander McLeod

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Physics and Astronomy

University College London

October 14, 2018

I, Michael Alexander McLeod, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

In the first part we shall consider the statistical analysis of large scale structure in galaxy surveys. We demonstrate a method for jointly constraining cosmology and photometric redshift distributions using cross correlations between photometric and spectroscopic redshift bins. This allows one to reduce the bias in the inferred cosmological parameters which may be propagated from errors in the redshift distributions. We demonstrate this using parameters for a DES-like survey, using galaxy number count $C(l)$s and CMB-TT information. We continue in this vein to apply these methods to the search for modified gravity using the Euclid survey. We forecast constraints on Horndeski theories using α-function parameterisation, using combined probes of galaxy number counts cross correlated with weak lensing shear, and independently adding CMB-TT information. We see that, as expected, the constraints on the α-parameters are not significantly degenerate with the other cosmological parameters; this is promising as it means that their detection would be less prone to misconstruction.

In the second part we consider the universe on considerably smaller scales, and concern ourselves with the local group (LG). We first explore the use of artificial neural networks for estimating the mass of the LG. Using the Timing Argument as a bench mark, we find that the ANN can make use of novel physical information (in our case, the eigenvalues and eigenvectors of the velocity shear tensor) to improve the scatter of the estimates considerably. We then proceed to explore the analytic Timing Argument mass further, exploring its dependency on dark energy and modified gravity models. Beginning with Λ, we proceed to perfect fluid models, quintessence fields, and scalar-tensor theories of gravity in the weak field limit.

Impact Statement

The work presented in Part I of this thesis is primarily dedicated to analysing and improving the constraining power of galaxy surveys in cosmology. The methods described for photometric redshift calibration can be implemented into pipelines for large collaborations, which will bring us our most precise and accurate measures of the universe. I hope that the work presented in the following pages can contribute to this deepening of our understanding of the universe. The work here also further explores a relatively recent approach to parameterised modified gravity, for the case of a major upcoming experiment (Euclid). The fact that we find ourselves able to constrain such a model, despite its additional complexities compared to the standard approach in modified gravity, shows that we can indeed use more intricate modelling in the near future. Our work also urges some caution, and I hope that our analysis of the stability of modified gravity models in the wake of gravitational waves can stimulate discussion on the future of parameterised modified gravity. In Part II we look to the universe directly around us, and find it a vibrant playground for physics and numerical methods of a wide variety. The solutions for simple, two body interactions in modified gravity is little explored in the literature, and this work could help bring the interesting effects of these theories to a different audience. The Local Group is typically considered in the context of very standard physics, the assumption being that the scales are too small and the physics too Newtonian for other models to be worth the hassle. The work presented here can show that even on scales as small as our nearest galaxy, we should be considering the full breadth of physics available to us. We present highly competitive estimates for the mass of the Local Group, which is a fundamental measurement of our own home. By remaining close to home, I hope that this work can be inspiring both inside and outside academia. I believe that public engagement with science is crucial both to us as scientists, and to society at large. As astronomers, we are in a particularly advantageous position to engage with the public's particularly fervent fascination with the space and the universe. I hope that the concrete example of our own galaxy and our closest partner (Andromeda) will be a gateway for the public into learning

about methods and ideas as complex and diverse as machine learning, galaxy flows, modified gravity, cosmological inference, and how we fit into the universe.

Previously Published Work

The work in chapter two is the basis of a paper published in MNRAS Volume 466, pp3358-3568, April 2017, entitled ‘A joint analysis for cosmology and photometric redshift calibration using cross-correlations’.

The work in chapter four is the basis of a paper published in JCAP December 2017, entitled ’Estimating the mass of the Local Group using machine learning applied to numerical simulations’.

Acknowledgements

It is a challenge in itself to thank everyone who has made this PhD possible.

I am, of course, indebted to my supervisors Dr. Filipe Abdalla and Prof. Ofer Lahav, who have guided me with their advice and expertise. They lead me to new questions I hadn't previously considered, that allowed me to grow as a person and a scientist.

Most importantly I wish to thank my parents, Joan McKenzie and David McLeod, for providing me extraordinary support throughout my entire education, and life in general. In particular I would like to thank my mother, Joan, for fostering my interest in the mathematical world from the earliest age, and thoroughly instilling in me the belief that education and intellectual exploration is a worthy cause for its own sake. In this way (and many others) she has probably influenced me more than any other person.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This is it, now soar my friend,
Up into your world of levity,
While I stay grounded here,
Underneath the thumb of gravity

The Solemn Prince

Throughout history our intense fascination with the cosmos has never waned; compelled to look upwards, we have longed to understand the planets, stars, galaxies, and all that came within our ever sharpening view. The story of astronomy and, later, physical cosmology, is really the story of gravity. The exploration of gravity became a field in its own right, and its formulation stands on very different grounds to the other three 'fundamental forces' identified in the current framework of physics. Their merger is no mean feat; quantum field theory and general relativity do not play nice. This is the first of the two great problems standing before our theories of gravity, and also the least likely to be solved any time soon. The other, though in many ways dating back to the infancy of general relativity (GR), appeared in its modern form quite by surprise in the 1990s.

Analysis of supernova data in two landmark papers [1][2] revealed for the first time that the expansion of the universe was not slowing down, as would be expected due to gravitational attractions, but in fact *accelerating*. Despite some tentative prior glimpses of a low mass universe from works such as [3], to find the result confirmed definitively was radical, and has since been corroborated many times over

by every cosmological probe conceivable. This was the rebirth of the cosmological constant as a major figure in physics, invoked to produce quite a different effect than that for which Einstein first conjured it. Over the following years, the current standard model of cosmology – the ΛCDM model – became established as the new scientific orthodoxy. Despite its success, Λ is by no means the only mathematical or scientific entity which could produce the effects which we are seeking to explain, and over the past two decades many people have found it unsatisfactory for one reason or another. Attempts have been made to identify the cosmological constant as a vacuum energy, but after producing an estimate often quoted as being some 120 orders of magnitude too large, one might say they have not been entirely successful. Problems with Λ were being identified as early as the 1980s [4], and although there has been much discussion on the subject, the question of the suitability of Λ is far from settled [6].

Instead of relying on Λ, we can widen our focus and look at the elusive *Dark Energy* (DE). Named in kind with the equally mysterious "Dark Matter" (whose name makes rather more sense), *Dark Energy* simply refers to whatever is driving the acceleration of the universe. This may take the form of a constant energy density, as in the case of vacuum energy or a cosmological constant, or it may take more dynamic forms such as perfect fluids or scalar fields. One class of theories is often separated from the label 'Dark Energy', perceived to be a different approach (although mathematically closely related), and is called *Modified Gravity* (MG). This class of theories, generally less widely supported than DE, proposes that, instead of the universe being dominated by some unknown (possibly highly exotic) source of energy density, GR may not be the whole story on cosmological scales. Instead, the theory is modified in some way (hence the rather straightforward title, though said modifications are no less shadowy in their own right) to produce some kind of repulsion on large scales.

The reason that these theories generate so much trouble and strife, and the reason we put up with it, is because they represent so much about what we don't know. It is astounding to comprehend that, according to most standard analyses,

some 96% of the universe is 'dark' i.e. unknown to us [76]. Just when we thought we were on the cusp of sussing out the universe it runs away from us. When in the history of physics have we had so many viable physical theories to model what we see? How can we possible explore such a vast and complex theory space? Do we stop at a single scalar field, as in the most basic DE and MG theories?

As our views opened up and stretched across the universe, we were forced to develop new ways to understand what we saw. Astronomy – in particular cosmology – became highly statistical. We now look at snapshots of millions or even billions of galaxies to calculate our quantities of interest, with time dependence only unfolding in the redshift evolution of the statistics. This approach has allowed us to observe the large scale behaviour of the universe, and check it against models of surprising simplicity. The evolution of the background and the bulk of the matter spectrum can be understood from just a few simple equations and less than a dozen parameters. Despite our ignorance of the 'dark universe', the standard model of cosmology has been highly successful. If only it were the only one.

1.1 An Overview of Gravity and Cosmology

Our theories of cosmology are certainly dominated by gravity – whatever form that might take. At very early times a mingling of poorly understood small scale forces were certainly important. The chronology of this very early universe is still largely speculative, dominated by as yet untested notions of unified forces, quantum gravity, and other very high energy effects. Things cool down after a supposed period of inflation (the evidence for which will be discussed a little later), and physics starts to look more familiar. Big Bang Nucleosynthesis (BBN) produces the first light elements, but as things cool down further and we settle into the gentle sweep of the 'Hubble flow', the behaviour of the universe on large scales is dominated by gravity. Over-densities pulled surrounding matter towards themselves, collapsing into structures which over time formed the magnificent details we see around us on every scale. Of course the standard model interactions play a significant role, without which we would of course have no cosmic microwave background (CMB)

radiation nor baryon acoustic oscillations (BAO), and the impact on the behaviour of baryonic matter (particularly in producing feedback mechanisms in active galactic nuclei) is hotly debated, but for the most part cosmology is the great gravitational observational experiment. Pity we only get one shot.

The standard model of gravity is general relativity (GR). A revolutionary theory which changed not only the way we view gravitational interactions but also space and time, it has proved itself remarkably successful at length scales spanning many orders of magnitude. First in solar system tests, it is now the basis of our satellite communications (and thus the reason why you can google the reason why that is), and produces our standard model of cosmology, which has produced results with startling precision.

1.1.1 The Chronology of the Universe

In this section, we will present a brief summary of the evolution of the universe, as suggested by the current standard models in physics. This is largely based on the chronologies in [25][29].

1.1.1.1 Up to 10^{-35}s: 'Big Bang' and Speculation

At this time the universe is extremely hot, dense, and mysterious to us. Prevailing opinion suggests that physical interactions are dominated by quantum gravity [20] [21] and an (entirely speculative) unified force [19]. Even 'big bang' itself is not really an event, since we have nothing with which to describe it or suggest that it really exists (it, of course, being a point which is mathematically undefined and therefore outside of physical models).

1.1.1.2 Inflation

Inflation is a mechanism introduced into cosmology by Guth in [9] in order to solve the 'horizon problem', the 'flatness problem', and the 'monopole problem'. Though hard evidence for inflation is still hard to come by, it is generally accepted as part of the standard model of cosmology, and provides the best source so far for the origins of cosmic structure formation. The basic tenet of inflation is that there is some field (or fields) called the inflation field or 'inflaton', which has an equation of state with

negative pressure i.e. $w < -\frac{1}{3}$. If the dynamics of space are dominated by this field, it leads to an exponential expansion, which is stopped under some conditions. The seeds of structure formation are sown by quantum fluctuations of this field, which lead to curvature fluctuations as the universe expands. These then correspond to matter over/under-densities which evolve through gravitational interactions. The fact that inflation predicts a roughly flat power spectrum of initial perturbations is probably the best evidence in support of it. Inflation also produces primordial gravitational waves, which ought to leave an imprint of the CMB polarisation. Searches for this signal have so far failed, but the BICEP3/Keck Array [18] experiment is being planned to try to find it.

1.1.1.3 Big Bang Nucleosynthesis

A few minutes into the life of the universe as we know it, things get cool enough for nuclear processes to start occurring. This creates the light elements – Hydrogen, Helium, and their isotopes. At this stage the universe is still hot enough for these to be in an ionised state, and the universe is thus filled with a hot, dense plasma. The abundance of species such as Deuterium can provide in particular a measure of the baryon density of the universe, as well as information about neutrinos and even physics beyond the standard model.

1.1.1.4 Recombination

As energies fall ever lower ($T \sim 1000K$, [25]), electrons and nuclei combine to form neutral atoms. This is a key moment in the history of the universe, because this is the moment that the universe becomes 'transparent'. Photons, which previously interacted very strongly with the electrically charged constituents of the plasma which dominated the universe, can now travel relatively unimpeded through the recently formed neutral matter fluid. In fact, many photons from this time still reach us today, redshifted to much larger wavelengths, in the form of the cosmic microwave background (CMB) radiation. The epoch of recombination defines the 'surface of last scattering', which is the last scattering processes to occur before the universe becomes transparent.

1.1.1.5 Reionisation

The epoch of reionisation is generally placed around $z \sim 6 - 15$ [7]. With atoms having formed, familiar structures can begin to form, such as stars. Besides synthesising heavy elements for the first time, they unleash ionising radiation into the universe, creating bubbles of ionised hydrogen around them which percolate throughout the universe. This ionisation does not interfere so greatly that it prevents our seeing the CMB, since the densities of matter are much lower today than in the past. Nevertheless, it gives us an observational window into the universe from HI absorption lines.

1.1.1.6 Galaxy Formation

Galaxies start to form at $\mathscr{O}(1\,\mathrm{GYr})$, but may be as early as 600 Myr [8]; these are a biased tracer of the matter density field. It has been suggested that in the early universe the number counts of galaxies was in fact rather high, due to their small size, and despite continued galaxy formation diminished due to successive mergers to form the larger galaxies that we see today.

1.1.1.7 Dark Energy Domination

As the universe expands the density of radiation and matter falls, whilst the density of dark energy remains (approximately, at least) constant. Thus at late times dark energy becomes the dominant component in the composition of the universe, and drives a further accelerated expansion (similar to inflation, but rather less extreme for now). The fractional density of DE surpasses matter at about $z \lesssim 0.5$, although DE become a significant component much earlier, and most studies of DE focus on data which is sensitive to dynamics in the range $z \lesssim 3$. This effect was only relatively recently discovered, and accounting for it has been one of the primary focuses of cosmology for the past twenty years.

1.1.2 There and Back Again: Newtonian Dynamics to General Relativity to the Weak Field Limit

Newtonian gravitation was certainly a major milestone in the development of physics: for the first time, we could understand the motion of celestial bodies in

relation to their intrinsic physical properties (cf. Kepler's purely geometric work). Or so we thought: Mercury had a pesky perihelion precession, and it was larger than expected. In fact, there seemed to be no satisfactory way of accounting for it in Newtonian mechanics. This, however, was not the motivation for GR. That theory came from much loftier origins, and solved a host of problems once it was fully realised.

1.1.2.1 Principles of General Relativity and the Einstein Field Equations

Gravity, as described by GR, takes the form of distortions to background space-time; it is not really viewed as a 'force' in the traditional sense at all. The source of this distortion is energy-density, which includes mass but also depends on other forms of energy such as radiation and the energy density of electro-magnetic fields. The Einstein field equations relate the geometry of a space-time to its contents. These closely coupled differential equations are, unsurprisingly, challenging to solve in many cases, but there are a number of simple scenarios with solutions which are relatively straightforward to derive. This includes, thankfully, the basis of our understanding of the expansion of a homogeneous and isotropic universe.

The geometry of the space-time is defined by the metric tensor $g_{\mu\nu}$, which may be defined implicitly by the line element

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu\,. \tag{1.1}$$

We will follow the convention of writing the Minkowski metric (representing flat space) as $\eta_{\mu\nu}$, and we use a $(+,-,-,-)$ sign convention such that

$$ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu = c^2\, dt^2 - dx^2 - dy^2 - dz^2\,. \tag{1.2}$$

There is always a frame of reference such that the tangent space at a given point P is Minkowski, i.e. a coordinate system may be found such that the metric in the local

neighbourhood may be expressed as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{1.3}$$

where $h \ll 1$, and the metric is exactly Minkowski at P. Particles which are freely falling (not subject to other forces) follow geodesics in the curved spacetime where $\frac{d\vec{p}}{d\tau} = \vec{0}$.

The Einstein field equations are concisely expressed in a single tensor equation, which belies their complexity:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -\kappa T_{\mu\nu}, \tag{1.4}$$

where $\kappa = \frac{8\pi G}{c^4}$. The quantity $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$ is sometimes called the Einstein tensor, $G_{\mu\nu}$, although it is often more sensible to leave it written explicitly to see the geometric dependencies. The Ricci scalar R is the contraction of the Ricci tensor $R_{\mu\nu}$, itself the contraction of the rank-4 curvature tensor $R_{\mu\nu\sigma\rho}$, which is furthermore defined in terms of the metric and the affine connections. So from the very simple looking equation – perhaps written simplest as $G_{\mu\nu} = \kappa T_{\mu\nu}$ – we can see a wealth of complexity as we unravel it! The energy momentum tensor, $T_{\mu\nu}$, will be defined for the purposes of cosmology typically as $\rho u_\mu u_\nu$, although we shall see in the variational approach to GR that it is assumed to take a particular form with respect to the standard model of physics.

Dark Energy may enter into the Einstein field equations in a number of ways. Some of this is simply down to interpretation: if I place a term on the left hand side it tends to be interpreted geometrically, and on the right as part of the energy-momentum. Mathematically this distinction is irrelevant. The simplest form of DE comes from a cosmological constant, which actually appears very naturally in the derivation of the Einstein field equations. This yields the equations

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -\kappa T_{\mu\nu} - \Lambda g_{\mu\nu}, \tag{1.5}$$

where I have placed the constant alongside the matter for no better reason than to make the equation look aesthetically more balanced. Based on the current standard model of cosmology, which includes Λ, this is the form of the Einstein field equations that one ought to make use of in general, and we shall use it wherever we are not employing a different model of DE. (In other words, we shall not be using the Einstein field equations with *no* DE, even in cases where this is often done. We shall see later that DE can be important even at relatively small scales.)

1.1.2.2 A Variational Approach to Gravitation

The theory of GR may also be derived from a variational principle [26]. There are a number of advantages to using a variational approach. Much of theoretical physics is based on the principle of least action, and thus having an appropriate action for GR allows us to combine it with other areas of physics. It also allows us to make modifications to the theory easily and in ways that are straight forwardly interpretable in terms of classical field theories. (We shall see that this is very helpful in theories of inflation, DE, and MG.)

In this case the action, known as the Einstein-Hilbert action is:

$$S^{EH} = \int d^4x \sqrt{-g}\,[R + \mathscr{L}^m]\,. \tag{1.6}$$

Given that $\sqrt{-g}\,d^4x$ is the volume element, this is just about the simplest action that we could write for a theory that involves R and a matter Lagrangian! The coupling between matter and the curvature is minimal; they are coupled by the term $\sqrt{-g}$ in the volume element and by terms of the form $g^{\mu\nu}\partial_\mu\psi\partial_\nu\psi$ in the matter Lagrangian (although in cosmology the matter Lagrangian is rarely specified explicitly, since one assumes it lies in the realm of QFT). In principle the matter Lagrangian is just the standard model on a curved background, although this is rarely if ever used in practice. In cosmological applications we almost always work with $T^{\mu\nu}$ directly in the field equations for standard matter fields (baryonic or dark). When dealing with gravitational theories there are in fact two different variational approaches that are widely used: metric variation, and metric-affine variation (also

known as Palatini variation) [27]. In the latter, the affine connection, $\Gamma^a_{\ bc}$, is allowed to vary independently of the metric. Both will yield the Einstein field equations using an action which is equivalent to equation 1.6 modulo total derivatives. (Metric and metric-affine variation do however produce different field equations in e.g. $f(R)$ theories. [28])

1.1.2.3 The Weak Field Limit

Once we have the gravitational field equations we need to be certain that they are consistent with Newtonian mechanics in the appropriate limit, given that Newtonian gravity has been well confirmed by a multitude of observational and experimental evidence. We very briefly summarise here the recovery of Newtonian physics in the weak field limit of GR. This will be useful later when we work in extensions to the standard model, and when we discuss the dynamics of the Local Group (LG), which is on scales small enough to be approximately Newtonian. The domain of application of Newtonian physics is weak gravitational fields and small velocities ($v \ll c$); this is not as restrictive as it seems, and much of cosmology and n-body simulations is based on predominantly Newtonian force equations. In fact, we will not recover the Newtonian law of gravitation, but a Newtonian law *with* Λ, since we will start from the Einstein equations with a cosmological constant. We will then observe that there is a further limit in which DE is not important and gravity is well described as Newtonian after all. The key assumptions that we must make are as follows:

- $\vec{u} \approx (c,0,0,0)$; this represents matter being slow moving.
- $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ for $h_{\mu\nu} \ll 1$; this represents weak gravitational perturbations.

In this case, the Einstein field equations are dominated by their 00 (or tt) component. [26] This leaves a single equation:

$$R_{00} = -\kappa\left(T_{00} - \frac{1}{2}T g_{00}\right) + \Lambda g_{00}. \tag{1.7}$$

The term R_{00} may be written in terms of the metric (after some simplification) as $R_{00} \approx -\frac{1}{2}\delta^{ij}\partial_i\partial_j h_{00}$; the term $T_{00} \approx \rho c^2$ by our assumption of matter being close to at rest. This allows us to rewrite the above equation as

$$\vec{\nabla}^2 h_{00} = \kappa\rho c^2 - 2\Lambda = \frac{8\pi G}{c^2}\rho - 2\Lambda. \tag{1.8}$$

This is structurally the same as the Newtonian Poisson equation (plus a cosmological constant term), and allows us to identify the gravitational potential with the metric term $\Phi = \frac{1}{2}h_{00}c^2$. We can then write the Newtonian theory as

$$\vec{\nabla}^2\Phi = 4\pi G\rho - \Lambda c^2, \tag{1.9}$$

which may then be integrated to give the acceleration equation [80][117]:

$$\vec{a} = \left(-\frac{GM}{r^2} + \frac{\Lambda c^2}{3}r\right)\hat{\vec{r}}. \tag{1.10}$$

The repulsive acceleration term due to Λ is equivalent to a uniform expansion of the intervening space. We return to a standard Newtonian law whenever this term is negligible i.e. when we are looking at systems over short enough distances. For the solar system this is certainly the case, with any Λ expansion being entirely unobservable at present; on scales of $\mathcal{O}(1\,\mathrm{Mpc})$ and above however it can become significant.

1.1.3 Background Cosmology and ΛCDM

Getting from GR to cosmology is not as large a step as it may seem. Motivated by observations of homogeneity and isotropy on the largest scales, the background solution that we use in almost all forms of cosmology (including theories with DE and MG) is the maximally symmetric solution to the Einstein field equations. This metric is know as the FLRW metric, and it looks like this:

$$ds^2 = c^2\,dt^2 - a(t)^2\left[\frac{dr^2}{1-kr^2} + r^2\,d\theta^2 + r^2\sin^2(\theta)\,d\phi^2\right]. \tag{1.11}$$

Where $k \in \{-1, 0, 1\}$, corresponding to the cases of open, flat, or closed universes respectively. In the case of a flat background spacetime (which we will assume throughout this thesis; Planck estimates of curvature are all consistent with zero [76], with Planck combined with BAO, supernova, and H_0 observations giving $\Omega_k = 0.001 \pm 0.004$) we have $k = 0$ and thus the metric reduces to the simple form

$$ds^2 = c^2\,dt^2 - a(t)^2\,d\vec{x}^2\,. \tag{1.12}$$

with the only deviation from our familiar Minkowski metric being a time dependent scale factor $a(t)$, which describes the expansion history of the universe. Co-moving (or fundamental) observers remain at fixed spatial coordinates, whilst the distances between co-moving objects scales with $a(t)$, and hence as the scale factor increases all co-moving objects recede from one another. Obviously, if everything in the universe were co-moving, then things would look very boring indeed. Structures can form because of peculiar motions relative to the expansion of the universe, and over-densities may become massive enough to overcome the expansion altogether, collapsing to form bound objects. Such objects move along with the expansion of the background, but do not themselves expand with it. Nevertheless, the background can be described by considering the mean densities of the various components of the universe, modelled as perfect fluids.

The expansion is governed by the Friedman equations, sometimes known as the cosmological field equations:

$$\ddot{a}(t) = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right)a(t) + \frac{1}{3}\Lambda c^2 a(t), \tag{1.13}$$

$$\dot{a}(t)^2 = \frac{8\pi G}{3}\rho a^2(t) + \frac{1}{3}\Lambda c^2 a^2(t) - \frac{kc^2}{R_0^2}. \tag{1.14}$$

Where, for a cosmology with multiple components,

$$\rho = \sum \rho_i, \tag{1.15}$$

$$\rho_i = a^{-3(1+w_i)}\rho_{i,0}, \tag{1.16}$$

$$p = \sum p_i, \tag{1.17}$$

$$p_i = w_i \rho_i c^2. \tag{1.18}$$

Here w_i is the equation of state of component i, which is defined implicitly by equation 1.18; ρ and p without subscripts refer to the total density and pressure respectively, whilst subscripts indicate the density or pressure of an individual component. Although we included the cosmological constant separately in this field equations, it is equivalent to a fluid with $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$ and $w_\Lambda = -1$ (which clearly leaves the density constant). We also have $w_m = 0$ for pressureless matter (a.k.a. dust, which serves as a model for non-interacting cold dark matter) and $w_r = \frac{1}{3}$ for radiation. Notice also that the curvature term includes R_0, which in a closed universe represents the radius of the universe at the present time. The normalised scale factor $a(t)$ is related by

$$a(t) = \frac{R(t)}{R_0}. \tag{1.19}$$

For the purposes of cosmological analysis, it is convenient to replace ρ_i with $\Omega_i = \frac{\rho_i}{\rho_{\text{crit}}}$,

$$\Omega_i = \frac{\rho_i}{\rho_{\text{crit}}} = \left(\frac{3H^2}{8\pi G}\right)^{-1} \rho_i, \tag{1.20}$$

where the critical density has been implicitly defined as $\rho_{\text{crit}} = \frac{8\pi G}{3H^2}$. The critical density is the mass / energy density required to maintain a flat universe; therefore in a flat universe all of the density contributions sum to the critical density and quantity Ω_i is the fractional density of that component in the universe. This is the form which we will give the parameters for cosmological inference later. They satisfy the closure relation

$$\sum \Omega_i = 1 - \Omega_k, \tag{1.21}$$

where Ω_k captures the curvature

$$\Omega_k = -\left(\frac{c}{RH}\right)^2 k. \tag{1.22}$$

One of the advantages of this is that, in a flat universe the dimensionless densities Ω_i represent the fraction of the energy density in the universe contributed by each sector.

1.1.4 The Cosmological Parameters

It is difficult to say exactly how many parameters exist in the ΛCDM model; people make different assumptions about which things are kept fixed or which details are modelled at all. We deal with at least these seven:

1. Ω_b, the fractional density of baryonic matter,
2. $\Omega_{\rm cdm}$, the fractional density of cold dark matter,
3. $\Omega_\Lambda/\Omega_{\rm DE}$, the fractional energy density of the cosmological constant or DE,
4. τ, the optical depth to reionisation,
5. A_s, the amplitude of (scalar mode) primordial curvature perturbations,
6. n_s, the spectral index,
7. w, the equation of state of DE ($\equiv -1$ for Λ).

These are enough to satisfy us, although there are a number of other parameters that may be of interest. In the following we shall discuss the parameters listed above as well as highlight some parameters of interest that we have chosen to leave out of our analyses.

1.1.4.1 Baryonic Matter Ω_b

The 'baryonic' matter is any matter consisting of standard model particles *not* including neutrinos. Baryonic matter is matter that is based on this (mostly Hydrogen, but also heavier nuclei and complex molecules). The initial properties of baryonic matter are determined by standard model interactions in the early universe, known as Big Bang Nucleosynthesis (BBN). In the early universe the baryonic matter is very hot (high energy), but as the universe expands it cools so that its energy is dominated by the rest mass energy. We say this matter is 'non-relativistic', because

$v \ll c$, and the dynamics (not including the background) are largely Newtonian. We further assume that this matter is almost pressureless, so we can approximate it as an ideal fluid with $p = 0$ (and hence $w = 0$). (This is an acceptable assumption as long as the matter has sufficiently low mean density that interactions are negligible.) Its most important imprint on the universe, besides being that which we are made of, are baryon acoustic oscillations (BAO) [15]. These acoustic oscillations originate from the tightly coupled baryon-photon fluid in the early universe (when matter was ionised and thus strongly interacting with photons); pressure waves in this fluid were frozen out at recombination when the two components ceased their interaction, and remained visible as density fluctuations. These are observable today in the power spectrum of the CMB or galaxies. Baryonic matter makes up approximately 5% of the energy density of the universe [76].

1.1.4.2 Cold Dark Matter Ω_{cdm}

The properties of cold dark matter (CDM) are quite basic in the form that it is normally used in cosmology. Firstly, it is non-interacting (making it pressureless) and hence its dynamics are determined entirely by gravitation (this is what makes it dark). Secondly, it is non-relativistic in the same sense as baryonic matter above (this is what makes it 'cold'). There are many theoretical proposals for dark matter (such a sterile neutrinos, axions, and WIMPs), with many possible interactions and particle masses, none of which have been favoured by any evidence as yet. We still have only upper bounds on interaction cross sections [14] (so non-interacting remains a good approximation) from either astrophysical or collider based experiments. CDM makes up some 25% of the energy density of the universe [76].

1.1.4.3 Dark Energy Ω_Λ, Ω_{fld}, Ω_φ ...

The density of dark energy (DE) dominates the 'energy budget' of the universe today, accounting for roughly 70% of the energy density of the universe. Despite this, it is still poorly understood. The simplest form is Λ, the cosmological constant, which has an effective equation of state $w = -1$. Λ is the most common form of DE in cosmology, used in the 'concordance model' ΛCDM.

The first step in a more adventurous direction is to introduce a perfect fluid

with a different equation of state. This equation of state may be a constant $w = w_0$, or it may vary over time, a common parameterisation being $w(a) = w_0 + w_a(1 - a)$. (Note that in either case w_0 is, by definition, the equation of state parameter for DE in the present.) This approach to DE is phenomenological, and seeks to parameterise the average evolution of w for a DE model.

A step further brings us to the realm of scalar fields, the simplest fully fledged DE (and inflation) models. We shall discuss these in more depth in the next section.

1.1.4.4 Optical depth at reionisation τ_r

Reionisation is a crucial epoch in the history of the universe. As dense structures form – stars and galaxies – energetic processes emit ionising radiation, which causes Hydrogen in the universe to re-ionise. Unlike in the very early universe, this Hydrogen is not dense enough to be optically opaque, so the CMB largely survives with some Thompson scattering. This scattering allows us to constrain reionisation using the CMB; it is almost impossible to study τ_r through large scale structure. Although we are focused on galaxy surveys, we will use CMB information to constrain certain parameters such as this one in this thesis.

1.1.4.5 The amplitude of fluctuations A_s

The primordial power spectrum (post-inflation) is taken as a starting point for most cosmology; the simplest form one usually takes is a power law form,

$$P(k) = A_s \left(\frac{k}{k^*} \right)^{n_s - 1}. \tag{1.23}$$

The scale of the spectrum is defined by an amplitude at some particular wavenumber k^*, known as the *pivot scale*. This scale may be chosen arbitrarily, and so in comparing results between studies, one must convert between amplitudes using:

$$A_{s,1} = \left(\frac{k_1^*}{k_2^*} \right)^{n_s - 1} A_{s,2}. \tag{1.24}$$

Due to its tiny size – $\mathcal{O}(10^{-9})$ – it is typically expressed in log-form for the purposes of computation.

1.1.4.6 The spectral index, n_s

The spectral index, n_s, is the exponent in the power law model of the power spectrum. One may express it as $\frac{d\log P}{d\log k}$. (Unlike A_s, it is not dependent on the pivot scale.) Inflation models generally suggest that n_s ought to be close to unity and data currently supports a value around 0.96; this is perhaps the greatest success of inflation.

There are a number of other parameters which we do not consider directly in this thesis, which are of interest to cosmologists, even within the standard model. These include:

- The tensor-scalar ratio, r: Inflation theories in general source primordial gravitational waves (tensor perturbations) as well as scalar curvature perturbations. The parameter r is the amplitude of tensor scalar perturbations relative to the scalar perturbations. Although BICEP2 originally claimed a detection [16], the data was later found to be consistent with $r = 0$ [17]. Detection of non-zero r would be powerful evidence for inflationary cosmology.

- Isocurvature perturbations: Isocurvature perturbations [25] are perturbations where the total energy density remains the same, but the balance of the components (e.g. matter and radiation) are different $\delta\rho_m = -\delta\rho_r$. That energy density remains constant implies that so does the curvature, and hence isocurvature modes. At the moment there is no evidence for isocurvature modes, which are strongly constrained by the CMB. Perturbations are generally taken to be adiabatic, where $\delta(1+w_i)^{-1}\rho_i = (1+w_j)^{-1}\rho_j$.

- Neutrinos: The presence of massive neutrinos has an impact on cosmology [10][11]. They can evidence themselves by the fact that they transition from being relativistic to non-relativistic at temperatures (times) characterised by their masses. The perturbations in the neutrino sector are also different, as their large velocities smooth out perturbations on scales less than their 'free streaming length'. Cosmology is rapidly becoming a competitive tool for neutrino physics. [12][13].

- Primordial non-Gaussianity f_{NL}: Non-Gaussianity is another prediction of some inflation theories [22]. If a field is Gaussian, then its statistics are fully described by the 2-point correlation functions; non-Gaussianity can be searched for by calculating the n-point correlation functions of the primordial power spectrum for $n > 2$.

1.2 Beyond ΛCDM: Dark Energy and Modified Gravity

The standard model of cosmology (often called the 'concordance model') has certainly become the ΛCDM model. Whilst the ΛCDM model is widely accepted, it is not unique in its ability to match observations (although it is the *simplest* model to do so successfully), and many other potential explanations have been sought. From the rotation curves of galaxies to the acceleration of the universe, as the principle force in play gravity has been a frequent target for modifications to fit the data. Although early alternative gravity theories such as Milgrom's modified Newtonian dynamics (MOND) [69] aimed to avoid the need for dark matter, modified gravity (MG) as discussed in this thesis (and generally in the literature) is taken to mean theories designed to account for the late time acceleration of the universe by modification of the Einstein-Hilbert action.

General relativity presents a concise theory based on a few intuitive principles which Einstein felt gravity should respect. In fact, it is about as concise as a theory can be which is geometric and respects the equivalence principle. It is, however, not the *only* theory that does so. We can modify the theory of gravity, either at the level of the action or at the level of the field equations, to create new theories of gravity designed to produce the accelerated expansion typically ascribed to dark energy or Λ. Here, however, we face an interesting problem. Much like with dark matter, only the broad characteristics of DE/MG are specified and there are simply too many options to choose from! [23] Detecting or ruling out MG would be an almost impossible – and certainly tedious – task if we were to consider each modification of gravity in turn. Even taking a single sub-theory, such as $f(R)$, leaves us with

a free function which cannot be constrained without parametrising in some simple way and restricting the theory. Indeed, most of these 'theories' are more *frameworks*, from which a given theory can be chosen and tested. If we wish to make some headway with MG, we need a way to consider a broad class of MG theories in a simple – and preferably physically intuitive – parameterisation. There are many ways of parameterising general modifications to gravity, and these parameters can then be included in standard Bayesian analyses. Often these are without physical intuition and therefore resist interpretation in any meaningful way. This would be a significant problem if one were to make a detection, since it would be very challenging to understand exactly *what* one had detected, and therefore to understand what changes to gravity we should be considering as part of a more complete theory at a fundamental level.

1.2.1 Scalar Fields in Cosmology: The Minimally Coupled Scalar Field

We begin with a brief introduction into scalar fields in cosmology, and how they can impact the dynamics of the universe. Scalar fields have been explored in the context of GR for many decades, with minimally coupled scalar fields being the basis for single field inflation [71] and quintessence [74]; this section is based primarily on modern reviews which may be found in [26] [29]. These theories retain Einstein's GR, in the sense of the Einstein-Hilbert part of the action remaining unchanged, with an additional minimally coupled scalar field (i.e. does not couple directly to the Ricci scalar or to the matter fields), and thus is usually considered a dark energy theory. In this section, we shall employ natural units in order to avoid algebraic clutter.

These theories are derived from an action, which looks like so:

$$S = \int d^4x \sqrt{-g} \left[R + \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right] + S^m(g^{\mu\nu}, \psi_i). \tag{1.25}$$

Where the metric dependence of the kinetic term has been made explicit. We

may vary the action with respect to the scalar field to find the scalar field equations of motion,

$$\Box\phi + V'(\phi) = 0. \tag{1.26}$$

For a free particle, this reduces to the familiar Klein-Gordon equation, although for our purposes we shall require a very different kind of potential! Since we have not modified the Einstein-Hilbert part of the action, the field equations for this theory are the same as before, with an additional component of energy-momentum tensor $T_{\mu\nu} = T^{\phi}_{\mu\nu} + T^{m}_{\mu\nu}$. This is calculated by varying the scalar field action with respect to the metric. The energy momentum tensor of the scalar field is given by

$$T^{\phi}_{\mu\nu} = \nabla_\mu\phi\nabla_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}\nabla_\sigma\phi\nabla^\sigma\phi - V(\phi)\right). \tag{1.27}$$

This equation can be compared to the energy momentum tensor of a perfect fluid:

$$T^{m}_{\mu\nu} = (\rho + p)u_\mu u_\nu - pg_{\mu\nu}. \tag{1.28}$$

If we compare these terms in the rest frame of the perfect fluid ($\mathbf{u} = (1,0,0,0)$), and in local Minkowski coordinates ($g_{\mu\nu} = \mathrm{diag}\{1,-1,-1,-1\}$), and finally with the assumption that the scalar field is not spatially varying (for the purposes of homogeneity), then we arrive at

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi), \tag{1.29}$$

$$p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{1.30}$$

So, just as with matter and radiation, we find ourselves with a perfect fluid, which may enter into the Friedman equations and contribute to the evolution of the universe. Unlike matter, radiation, or Λ, the equation of state parameter w_ϕ is not necessarily a constant, and can evolve in potentially complex ways:

$$w_\phi = \frac{p_\phi}{\rho_\phi} = \frac{\dot{\phi}^2 - 2V(\phi)}{\dot{\phi}^2 + 2V(\phi)}. \tag{1.31}$$

One may easily see that Λ is a special case of this theory, where $\dot{\phi} = 0$ and thus $w = -1$. An accelerated expansion can be achieved as long as $w_\phi < -\frac{1}{3}$.

1.2.2 Scalar-Tensor Actions in the Jordan and Einstein Frames

Scalar-tensor theories of gravity look similar to the minimally coupled scalar field theories discussed above, but they can behave very differently. This is because the scalar field is now non-minimally coupled to the geometry (usually via the Ricci scalar) or to matter. In fact, the metric or matter coupling depends on the frame of reference: a theory which is minimally coupled to matter is related to a theory which is minimally coupled to the geometry via a conformal transformation of the metric. (A conformal transformation is an angle preserving transformation which will be defined mathematically in the text that follows. It can be achieved by scaling the metric at every point by a strictly positive function of the spacetime.)

Probably the earliest scalar-tensor theory is Brans-Dicke [129] which has been constrained many times by solar system tests [154] and cosmology [130]. An excellent review of these theories is found in [29], although there is also an extensive literature of research and pedagogical papers [67][131]. (Notations and conventions vary considerably between sources.)

The frame in which the scalar field is coupled directly to the geometry (and minimally coupled to matter) is known as the Jordan frame. Although there is a more general form of the action of Scalar-Tensor theories (see Horndeski gravity, section 1.2.3), when considering explicit theories we will look at the simpler and more common form,

$$S = \int d^4x \sqrt{-g} \left[A(\xi)R + \frac{1}{2} g^{\mu\nu} (\partial_\mu \xi)(\partial_\nu \xi) - V(\xi) \right] + S^m(g^{\mu\nu}, \psi_i). \qquad (1.32)$$

Where $A(\xi)$ is an arbitrary function of the scalar field ξ. This action contains many well known models, such as symmetrons, chameleons, and even $f(R)$ gravity, which may be recast as a scalar field theory. In this frame, the field equations no longer take their usual form due to the coupling of the geometry to the scalar field. However, since the matter action S^m is the identical to the standard model form,

the energy momentum tensor keeps its familiar form (plus an additional scalar field component) and free particles will follow geodesics as usual, since there are no interaction terms between the matter fields ψ_i and the scalar field ξ.

Another useful frame for looking at scalar-tensor theories is the Einstein frame, in which a redefinition of the metric ($g_{\mu\nu} \to \tilde{g}_{\mu\nu}$) means that the scalar field coupling to $\tilde{R}$ vanishes, but the (redefined) scalar field ϕ appears in the matter action. The metrics are related by some conformal transformation

$$g_{\mu\nu} = \Omega^2(\phi(\mathbf{x}))\tilde{g}_{\mu\nu}, \tag{1.33}$$

where $\Omega^2 > 0$ everywhere. (For this reason, it is sometimes rewritten as an exponential $\Omega = e^{\omega(\phi)}$, which also makes Taylor expansions for small $\omega(\phi)$ simple.) We may likewise write its inverse metric

$$g^{\mu\nu} = \Omega^{-2}(\phi(\mathbf{x}))\tilde{g}^{\mu\nu}, \tag{1.34}$$

which must be true if we are to satisfy

$$g^{\mu\nu}g_{\mu\nu} = \tilde{g}^{\mu\nu}\tilde{g}_{\mu\nu} = \delta^{\mu}_{\nu}. \tag{1.35}$$

It may be worth noting that if $\phi = \phi(t)$ then the conformal transformation becomes just a time dependent scale factor. The conformal factor may be easily defined in terms of the Jordan frame action:

$$\Omega^2(\xi) = A(\xi). \tag{1.36}$$

If the transformation is anywhere singular, then the theory is physically ill-defined. The Ricci scalar transforms as

$$\tilde{R} = \frac{1}{A}\left[R - 3\nabla^2\log(A) - \frac{3}{2}\left(\nabla^{\mu}\log(A)\right)\left(\nabla_{\mu}\log(A)\right)\right]; \tag{1.37}$$

using these relations, we can rewrite the action using these new variables (omitting

total derivatives which do not contribute to the equations of motion).

$$S = \frac{1}{16\pi G}\int d^4x\,\sqrt{-\tilde{g}}A^{-2}\left[A^2\tilde{R} + \left(\frac{3}{2}A^2\left(\frac{\partial \log(A)}{\partial \xi}\right)^2 + A\right)\left(\frac{\partial \xi}{\partial \phi}\right)^2 \tilde{g}^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi) + \tilde{V}(\phi)\right]. \tag{1.38}$$

From this we may infer the field redefinition

$$\left(\frac{\partial \phi}{\partial \xi}\right)^2 = \left(\frac{3}{2}\left(\frac{\partial \log(A)}{\partial \xi}\right)^2 + \frac{1}{A}\right). \tag{1.39}$$

We may write the Einstein frame action to make the ϕ dependence of S^m explicit:

$$S = \int d^4x\,\sqrt{-\tilde{g}}\left[\tilde{R} + \frac{1}{2}\tilde{g}^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi) - \tilde{V}(\phi)\right] + S^m(\Omega^{-2}(\phi)\tilde{g}^{\mu\nu}, \psi_i). \tag{1.40}$$

This action will reproduce the Einstein field equations (since nothing is coupled to $\tilde{R}$ and hence the Einstein-Hilbert action remains intact), but dynamics in this frame are still non-trivial: the new matter coupling means that particles will no longer travel along geodesics anywhere $\Omega(\phi) \neq 1$. (Note that, since a conformal coupling leaves null geodesics unchanged, photons and other massless particles *do* follow their usual geodesic paths.) Effectively the coupling between the matter and the scalar field is creating an additional force which is pushing the particles off of the geodesics; particles are only 'free' when the conformal factor is unity – usually when the scalar field is zero – and hence the coupling terms disappear, in which case the two metrics are the same and particles will travel along geodesics.

Since the actions are equivalent, their dynamics must be equivalent. The redefinition of the metric has not changed the coordinates of any particles, so they still follow the same paths in coordinate space, but since we have redefined the metric in going from one frame into the other, the particles cannot possibly follow geodesics in both. We are left with a trade off: the simplicity of the field equations in the Einstein frame, or the simplicity of the equations of motion in the Jordan frame. Unlike the field theories discussed in the previous section, since this scalar field couples directly to either matter or the metric, we cannot take it to be spatially homogenous, but must in some way trace either the matter distribution or the curvature.

1.2.2.1 Coordinates and Quantities in the Einstein and Jordan Frames

Firstly, the conformal transformation is *not* a coordinate transformation, merely a redefinition of the distances between points. Therefore

$$x^\mu = \tilde{x}^\mu. \tag{1.41}$$

Distances and proper times however must be affected by the scaling of the metric tensor,

$$ds^2 = g_{\mu\nu}x^\mu x^\nu = \Omega^2 \tilde{g}_{\mu\nu}\tilde{x}^\mu \tilde{x}^\nu = \Omega^2 d\tilde{s}^2, \tag{1.42}$$

$$d\tau^2 = \Omega^2 \, d\tilde{\tau}^2, \tag{1.43}$$

$$d\tau = \Omega d\tilde{\tau}. \tag{1.44}$$

These then naturally lead us to the following conclusions:

$$x_\mu = g_{\mu\nu}x^\nu = \Omega^2 \tilde{g}_{\mu\nu}\tilde{x}^\nu = \Omega^2 \tilde{x}_\mu, \tag{1.45}$$

$$u^\mu = \frac{dx^\mu}{d\tau} = \frac{d\tilde{x}^\mu}{\Omega d\tilde{\tau}} = \Omega^{-1}\tilde{u}^\mu, \tag{1.46}$$

$$u_\mu = g_{\mu\nu}u^\nu = \Omega^2 \tilde{g}_{\mu\nu}\Omega^{-1}\tilde{u}^\nu = \Omega \tilde{u}_\mu. \tag{1.47}$$

So although we have not performed a coordinate transformation, the covariant components of coordinate vectors are scaled, as are proper times and proper 4-velocities. It is worth noting that, since $x^\mu = \tilde{x}^\nu$ the coordinates of any particles are the same in both frames, and furthermore

$$v^i = \frac{dx^i}{dt} = \frac{dx^i}{dx^0} = \frac{d\tilde{x}^i}{d\tilde{x}^0} = \tilde{v}^i, \tag{1.48}$$

so coordinate velocities are also identical between the two frames.

1.2.2.2 Klein-Gordon Equations for the Scalar Field

How do we find out how the scalar field behaves? Varying the action w.r.t. ϕ (or ξ) leads to the Euler-Lagrange equations for the Scalar field:

$$\frac{\partial \mathscr{L}}{\partial \phi} - \nabla_\mu \left[\frac{\partial \mathscr{L}}{\partial \nabla_\mu \phi} \right] = 0, \tag{1.49}$$

$$\Box^2 \xi + V'(\xi) - A'(\xi) R = 0, \tag{1.50}$$

in the Jordan frame, or in the Einstein frame:

$$\Box^2 \phi + \tilde{V}'(\phi) - \frac{1}{\sqrt{-\tilde{g}}} \frac{\partial \mathscr{L}^m}{\partial \phi} = 0. \tag{1.51}$$

We may make the equation more intuitive and useful in the Einstein frame using an argument with a few subtleties. Since Ω^2 does not depend on $\partial_\mu \phi$, then neither do $\mathscr{L}^m$ or $g^{\mu\nu}$, and therefore

$$\frac{\delta \mathscr{L}^m}{\delta \phi} = \frac{\partial \mathscr{L}^m}{\partial \phi}, \tag{1.52}$$

$$\frac{\delta g^{\mu\nu}}{\delta \phi} = \frac{\partial g^{\mu\nu}}{\partial \phi}. \tag{1.53}$$

(We bear in mind that $\tilde{g}_{\mu\nu}$ and ϕ are being treated as independent variables in the variation, and that thus $g_{\mu\nu}$ does indeed depend on ϕ but not on any derivatives.) One can differentiate the metric with respect to the scalar field using

$$\frac{\partial g^{\mu\nu}}{\partial \phi} = \frac{\partial (\Omega^{-2} \tilde{g}^{\mu\nu})}{\partial \phi} = -2 \frac{\Omega'}{\Omega^3} \tilde{g}^{\mu\nu} = -\frac{2\Omega'}{\Omega} (\Omega^{-2} \tilde{g}^{\mu\nu}) = -2 \frac{\Omega'}{\Omega} g^{\mu\nu}, \tag{1.54}$$

and

$$\frac{\partial \tilde{g}^{\mu\nu}}{\partial \phi} = \frac{\partial g^{\mu\nu}}{\partial \phi} \frac{\partial \tilde{g}^{\mu\nu}}{\partial g^{\mu\nu}} = -2 \frac{\Omega'}{\Omega} g^{\mu\nu} \Omega^2 = -2 \frac{\Omega'}{\Omega} \tilde{g}^{\mu\nu}. \tag{1.55}$$

We may therefore evaluate the matter Lagrangian term as

$$\frac{1}{\sqrt{-\tilde{g}}} \frac{\partial \mathscr{L}^m}{\partial \phi} = \frac{1}{\sqrt{-\tilde{g}}} \frac{\delta \mathscr{L}^m}{\delta \tilde{g}^{\mu\nu}} \frac{\delta \tilde{g}^{\mu\nu}}{\delta \phi} = -\frac{\Omega'}{\Omega} \tilde{g}^{\mu\nu} \tilde{T}^m_{\mu\nu} = -\frac{\Omega'}{\Omega} \tilde{T}^m, \tag{1.56}$$

and the Einstein frame KG equation as

$$\Box^2\phi + \tilde{V}'(\phi) + \frac{\Omega'(\phi)}{\Omega(\phi)}\tilde{T}^m = 0. \tag{1.57}$$

This allows us to calculate the value of the scalar field in the Einstein frame using the energy-momentum tensor with which we are familiar. It is also often written (equivalently)

$$\frac{\Omega'}{\Omega}\tilde{T}^m = \partial_\phi\left(\log\Omega\right)\tilde{T}^m, \tag{1.58}$$

which is particularly useful when we use the exponential form of Ω^2. Since modifications are generally small, especially in the weak field limit, we may assume $\Omega \approx 1$. In this case, $\log(\Omega) \approx \Omega - 1$ using the Taylor expansion. This allows us to write an effective potential

$$\tilde{V}_{\text{eff}}(\phi) = \tilde{V}(\phi) + \log\left(\Omega(\phi)\right)\tilde{T}^m \approx \tilde{V}(\phi) + \left(\Omega(\phi) - 1\right)\tilde{T}^m, \tag{1.59}$$

and thus a simplified form of the KG equation

$$\Box^2\phi + \frac{\partial\tilde{V}_{\text{eff}}}{\partial\phi} = 0. \tag{1.60}$$

1.2.2.3 Einstein Field Equations

The field equations in the Einstein frame are the same as in GR, with an energy-momentum tensor that takes into account the new components due to the scalar field,

$$\tilde{R}_{\mu\nu} + \tilde{R}\tilde{g}_{\mu\nu} = \kappa\tilde{T}_{\mu\nu}. \tag{1.61}$$

Let's take a look at $\tilde{T}$, and how it is related to what we know (the matter distribution T^m and the scalar field ϕ). To begin, we can divide the tensor into two parts,

$$\tilde{T}_{\mu\nu} = \tilde{T}^m_{\mu\nu} + \tilde{T}^\phi_{\mu\nu}, \tag{1.62}$$

with contributions from the matter action S^m and the scalar field (potential and kinetic terms) respectively. The matter component is, as expected,

$$\tilde{T}^m_{\mu\nu} = \frac{1}{\sqrt{-\tilde{g}}} \frac{\delta \mathscr{L}^m}{\delta \tilde{g}^{\mu\nu}}, \tag{1.63}$$

but given that

$$\tilde{g} = \det(\tilde{g}_{\mu\nu}) = \det(\Omega^{-2} g_{\mu\nu}) = \Omega^{-8} \det(g_{\mu\nu}) \implies \sqrt{-\tilde{g}} = \Omega^{-4} \sqrt{g}, \tag{1.64}$$

and furthermore

$$\frac{\delta \mathscr{L}^m}{\delta \tilde{g}^{\mu\nu}} = \frac{\delta \mathscr{L}^m}{\delta g^{\mu\nu}} \frac{\partial g_{\mu\nu}}{\partial \tilde{g}_{\mu\nu}} = \frac{\delta \mathscr{L}^m}{\delta g^{\mu\nu}} \Omega^{-2}, \tag{1.65}$$

we thus have

$$\tilde{T}^m_{\mu\nu} = \Omega^2 T^m_{\mu\nu}. \tag{1.66}$$

The matter component of the energy-momentum tensor is the same as that which we typically use, but with a scaling according to the conformal transformation. We may understand this better by looking at the equations for dust.

$$T_{\mu\nu} = \rho u_\mu u_\nu, \tag{1.67}$$

$$\tilde{T}_{\mu\nu} = \Omega^2 \rho u_\mu u_\nu, \tag{1.68}$$

$$\tilde{T}_{\mu\nu} = (\Omega^4 \rho) \tilde{u}_\mu \tilde{u}_\nu, \tag{1.69}$$

$$\implies \tilde{\rho} = \Omega^4 \rho. \tag{1.70}$$

This is the transformation to ρ that we expect, since the total energy density contained within fixed coordinate limits should be invariant under the transformation, i.e.

$$\int \tilde{\rho} \sqrt{-\tilde{g}}\, d^4x = \int (\Omega^4 \rho)(\Omega^{-4} \sqrt{-g})\, d^4x = \int \rho \sqrt{-g}\, d^4x. \tag{1.71}$$

The ϕ component is also familiar, given that it is the same as the energy-

momentum tensor for a free scalar field, as in scalar field inflation or quintessence,

$$\tilde{T}^{\phi}_{\mu\nu} = \nabla_\mu\phi\nabla_\nu\phi - \tilde{g}_{\mu\nu}\left[\frac{1}{2}\nabla_\sigma\phi\nabla^\sigma\phi - \tilde{V}(\phi)\right]. \tag{1.72}$$

Although we may solve the KG equations for ϕ, we must also know $\tilde{g}_{\mu\nu}$ or make some simplifying assumptions to proceed.

1.2.2.4 Newtonian Limits

The assumptions that we make in finding a non-relativistic, Newtonian limit are:

- Particles are slowly moving ($v \ll c$).
- $g_{\mu\nu} = 1 + h_{\mu\nu}$ and only terms to first order in $h_{\mu\nu}$ are important.
- Newtonian potential is determined by the h_{00} component.
- Matter and scalar field are slowly evolving so time derivatives are negligible compared to spatial derivatives ($\Box^2 \to -\nabla^2$).
- The effects of the scalar field are small in the sense that $\Omega \equiv e^{\omega(\phi)} \approx 1 + \omega(\phi)$ and we need only take terms to first order in $\omega(\phi)$.
- Matter is slowly moving so $u^\mu \approx \{c, 0, 0, 0\}$.

With these assumptions in place we can derive an approximate Newtonian style potential. Following the typical procedure for GR we note that

$$\tilde{R}_{00} = -\kappa(\tilde{T}_{00} - \frac{1}{2}\tilde{T}\tilde{g}_{00}), \tag{1.73}$$

$$\tilde{R}_{00} \approx -\partial_i\tilde{\Gamma}^i_{00}, \tag{1.74}$$

$$\tilde{\Gamma}^i_{00} \approx \frac{1}{2}\delta^{ij}\partial_j\tilde{h}_{00}. \tag{1.75}$$

The first statement is just the 00-component of the field equations, and the next two are purely geometric statements assuming only that perturbations to Minkowski

space are small, and thus all of these should remain totally valid. Continuing then,

$$\tilde{R}_{00} \approx -\frac{1}{2}\delta^{ij}\partial_i\partial_j\tilde{h}_{00}, \tag{1.76}$$

$$\vec{\nabla}^2\tilde{h}_{00} \approx 2\kappa(\tilde{T}_{00} - \frac{1}{2}\tilde{T}). \tag{1.77}$$

You may notice that here we have dropped a term $\kappa\tilde{T}\tilde{h}_{00}$. Since $\tilde{h}_{00}$ is $O(\tilde{T})$, it follows that this term is $O(\tilde{h}^2)$ and so can be ignored. To proceed we must calculate $\tilde{T}_{00}$,

$$\tilde{T}^{\phi}_{00} = (\partial_t\phi)^2 - \tilde{g}_{00}\left[\frac{1}{2}\nabla_\sigma\phi\nabla^\sigma\phi - V(\phi)\right]. \tag{1.78}$$

Assuming that the scalar field is slowly evolving, then we may ignore the first term. We are left with a kinetic term and a potential term for the scalar field.

$$\tilde{T}^{\phi}_{00} = V(\phi) - \frac{1}{2}(\nabla\phi)^2, \tag{1.79}$$

$$\tilde{T}^{\phi} \approx (\nabla\phi)^2 + \eta^{\mu}_{\mu}\left[V(\phi) - \frac{1}{2}(\nabla\phi)^2\right] = 4V(\phi) - (\nabla\phi)^2. \tag{1.80}$$

1.2.2.5 Einstein frame weak field metric:

$$\tilde{T}^{m}_{00} = \tilde{\rho}c^2, \tag{1.81}$$

$$\tilde{T}^{m} = \tilde{\rho}c^2. \tag{1.82}$$

This gives us in the end

$$\vec{\nabla}^2\tilde{h}_{00} \approx 2\kappa\left(\frac{1}{2}\tilde{\rho}c^2 + V(\phi) - \frac{1}{2}(\nabla\phi)^2 - \frac{1}{2}\left(-(\nabla\phi)^2 + 4V(\phi)\right)\right), \tag{1.83}$$

$$\vec{\nabla}^2\tilde{h}_{00} \approx \kappa\left(\tilde{\rho}c^2 - V(\phi)\right). \tag{1.84}$$

Compared to GR,

$$\vec{\nabla}^2 h^{GR}_{00} \approx \kappa\rho c^2, \tag{1.85}$$

we have a modification to the metric from the gradient of the scalar field, but it is

otherwise similar to the Newtonian potential. Notice that it is only the same as the Newtonian potential if the gradient of the field is negligible, which means that the energy-momentum of the scalar field is negligible compared to matter and that it is slowly spatially varying.

1.2.2.6 The Newtonian Klein-Gordon Equation

The KG equation also simplifies in the Newtonian limit. Ignoring time derivatives we have $\partial_\mu \partial^\mu \to -\vec{\nabla}^2$; furthermore $T^m = \rho c^2$ and $\partial_\phi \log \Omega = \omega'(\phi)$. Putting this together we arrive at the Einstein frame Newtonian KG equation,

$$\vec{\nabla}^2 \phi + \frac{1}{2}\left(\omega'(\phi)\rho_m c^2 + V'(\phi)\right) = 0. \tag{1.86}$$

1.2.2.7 Motions of Test Bodies

The motion of test bodies may be calculated in either the Einstein or the Jordan frame. In the Jordan frame, we have already seen that the matter energy-momentum tensor is divergenceless,

$$\nabla_\mu T_m^{\mu\nu} = 0, \tag{1.87}$$

and therefore that dust follows geodesics [26]. The geometry however is more complex than in the GR case, because it is influenced by the scalar field. It is usually easier to look at dynamics of scalar field theories in the Einstein frame; we shall then convert to the Jordan frame to confirm that these motions correspond to Jordan frame geodesics.

1.2.2.8 Equations of Motion from $\nabla_\mu T^{\mu\nu}$

The conservation equation,

$$\nabla_\mu T^{\mu\nu} = 0, \tag{1.88}$$

must – and does – still hold in modified gravity theories. It is often said that this is violated in scalar-tensor theories in the Einstein frame but this is a misinterpretation: the equation is violated if we consider *matter only*. Thus, whilst $\nabla_\mu T_{\rm m}^{\mu\nu} \neq 0$ in general, we *do* have

$$\nabla_\mu \left[T_{\rm m}^{\mu\nu} + T_\phi^{\mu\nu}\right] = 0 \tag{1.89}$$

for a scalar-tensor theory. The energy-momentum tensor for a scalar field is given by

$$T_\phi^{\mu\nu} = \nabla^\mu\phi\nabla^\nu\phi - g^{\mu\nu}\left[\frac{1}{2}\nabla_\sigma\phi\nabla^\sigma\phi - V(\phi)\right]. \tag{1.90}$$

Taking the divergence we have

$$\nabla_\mu T_\phi^{\mu\nu} = (\nabla_\mu\nabla^\mu\phi)\nabla^\nu\phi + \nabla^\mu(\nabla_\mu\nabla^\nu\phi) - (\nabla^\nu\nabla_\sigma\phi)\nabla^\sigma\phi + \nabla^\nu\phi\frac{\partial V}{\partial\phi} \tag{1.91}$$

$$= \left(\Box^2\phi + \frac{\partial V}{\partial\phi}\right)\nabla^\nu\phi. \tag{1.92}$$

For a free scalar field, $T_\phi^{\mu\nu} = 0$ by virtue of the Klein-Gordon equation (and thus the scalar field and the matter fields are independently divergence-free). However, for a coupled scalar field the free field equation is modified by a matter term. We may still relate equation 1.91 it to the KG equation by

$$\Box^2\phi + \tilde{V}'(\phi) = -\partial_\phi\log(\Omega)\tilde{T}^m, \tag{1.93}$$

$$\implies \nabla_\mu T_\phi^{\mu\nu} = (\nabla^\nu\log(\Omega))\tilde{T}^m. \tag{1.94}$$

Now turning our attention to the matter sector, we may write (for dust)

$$T_m^{\mu\nu} = \rho u^\mu u^\nu. \tag{1.95}$$

Taking the divergence,

$$\nabla_\mu T^{\mu\nu} = \rho(\nabla_\mu u^\mu)u^\nu + \rho u^\mu(\nabla_\mu u^\nu) - \rho c^2(\nabla^\nu\log(\Omega)) = 0, \tag{1.96}$$

and contracting this expression with u_ν, we have

$$\rho(\nabla_\mu u^\mu)c^2 + \rho u^\mu(\nabla_\mu u^\nu)u_\nu - \rho c^2(\nabla^\nu\log(\Omega))u_\nu = 0. \tag{1.97}$$

The second term is identically zero since $u^\nu u_\nu = c^2$. Given the approximation that $u_\nu \approx (c,0,0,0)$, the third term is $\approx \rho c^2\partial_{ct}\log(\Omega)c \approx 0$ since we are ignoring time

derivatives. As such, we obtain as usual

$$\nabla_\mu u^\mu \approx 0. \tag{1.98}$$

We plug this back into our expression for the divergence to find

$$u^\mu(\nabla_\mu u^\nu) - c^2(\nabla^\nu \log(\Omega)) = 0. \tag{1.99}$$

The first term is a convective derivative and shows the modification to geodesic motion,

$$\ddot{x}^\nu + \Gamma^\nu_{00}\left(\frac{dt}{d\tau}\right)^2 - (\nabla^\nu \log(\Omega)) = 0. \tag{1.100}$$

Using in the Newtonian approximation $\Gamma^0_{00} \approx 0$, and $\Gamma^i_{00} = \frac{1}{2}\delta^{ij}\partial_j h_{00}$, and using $\nabla^i \approx -\nabla_i$,

$$\frac{d^2x^i}{dt^2} \approx -\partial_i\left(\frac{\tilde{h}_{00}}{2} + \log\Omega\right). \tag{1.101}$$

But above we saw that h_{00} also depends on the potential $\tilde{V}(\phi)$. For a constant potential this would be a give Λ like effect in addition to the Newtonian potential, and then there would be a further correction due to the matter coupling Ω.

We may convert to the Jordan frame using the relations in section 1.2.2.1, provided that any assumptions made are carried over between frames. Converting just equation 1.101, the only quantity that changes is h_{00}. Since perturbations to the metric are small $\Omega \approx 1 + \log\Omega$, and thus to first order

$$h_{00} \approx \tilde{h}_{00} + 2\log\Omega. \tag{1.102}$$

Inserting this into equation 1.101 yields

$$\frac{d^2x^i}{dt^2} \approx -\partial_i\left(\frac{h_{00}}{2}\right), \tag{1.103}$$

which is just the standard geodesic motion equation in the weak field, quasi-static limit.

1.2.3 The Horndeski Class

The Horndeski class is a subclass of scalar-tensor MG theories, which includes just one scalar field φ with at most second order equations of motion. It was first derived in [133], long before there was any interest in dark energy or modified gravity. (Indeed, the paper never even mentions gravity.) It was revived recently due to increasing interest in models such as covariant galileons [134] [135], which are not described by the action in the previous section, and a desire to classify theories for exploration. This class contains many models of interest and, hopefully, allows them to be simply parameterised and therefore marginalised over.

The general Lagrangian can be written

$$S = \int \sqrt{-g} \left[\sum_{i=2}^{5} \mathscr{L}_i + \mathscr{L}_m[g_{\mu\nu}] \right], \tag{1.104}$$

where there are four Lagrangian components $\mathscr{L}_i$ which characterise the dynamics and effects of the scalar field. (The rather unusual numbering beginning at 2 is historical.) Each of the Lagrangians relates to a different aspect of the theory.

$$\mathscr{L}_2 = K, \tag{1.105}$$

$$\mathscr{L}_3 = -G_3 \Box\varphi, \tag{1.106}$$

$$\mathscr{L}_4 = G_4 R + \frac{G_4}{dX} \left[(\Box\varphi)^2 - \nabla_\mu\nabla_\nu\varphi\nabla^\mu\nabla^\nu\varphi \right], \tag{1.107}$$

$$\mathscr{L}_5 = G_5 G_{\mu\nu}\nabla^\mu\nabla^\nu\varphi \tag{1.108}$$

$$- \frac{1}{6}\frac{dG_5}{dX} \left[(\Box\varphi)^3 + 2(\nabla_\mu\nabla^\nu\varphi)(\nabla_\nu\nabla^\alpha\varphi)(\nabla_\alpha\nabla^\mu\varphi) - 3\nabla_\mu\nabla_\nu\varphi\nabla^\mu\nabla^\nu\varphi\Box\varphi \right], \tag{1.109}$$

where K, and G_i are free functions of (ϕ, X) with $X \equiv -\frac{1}{2}\nabla^\mu\phi\nabla_\mu\phi$, and $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ is the usual Einstein tensor. The terms $\mathscr{L}_{2,3}$ involve only the scalar field, and hence will contain information such as the bare potential of the scalar field and the scalar field kinetic terms. The coupling to gravity, via R and $G_{\mu\nu}$, comes from $\mathscr{L}_{4,5}$.

The Lagrangian is expressed in terms of four free functions; the linear pertur-

bations may also be expressed in terms of four free functions of time $\alpha_i(t)$, which may be directly related to the functions in the Lagrangian. The α-functions are defined as follows [153],

$$M_*^2 = 2(G_4 - 2XG_{4X} + XG_{5\phi} - \dot{\phi}HXG_{5X}), \tag{1.110}$$

$$HM_*^2\alpha_M = \frac{dM_*^2}{dt}, \tag{1.111}$$

$$H^2M_*^2\alpha_K = 2X(K_X + 2XK_{XX} - 2G_{3\phi} - 2XG_{3\phi X}) \tag{1.112}$$

$$+ 12\dot{\phi}XH(G_{3X} + XG_{3XX} - 3G_{4\phi X} - 2XG_{4\phi XX}) \tag{1.113}$$

$$+ 12XH^2(G_{4X} + 8XG_{4XX} + 4X^2G_{4XXX}) \tag{1.114}$$

$$- 12XH^2(G_{5\phi} + 5XG_{5\phi X} + 2X^2G_{5\phi XX}) \tag{1.115}$$

$$+ 4\dot{\phi}XH^3(3G_{5X} + 7XG_{5XX} + 2X^2G_{5XXX}), \tag{1.116}$$

$$HM_*^2\alpha_B = 2\dot{\phi}(XG_{3X} - G_{4\phi} - 2XG_{4\phi X} \tag{1.117}$$

$$+ 8XH(G_{4X} + 2XG_{4XX} - G_5\phi - XG_{5\phi X}) \tag{1.118}$$

$$+ 2\dot{\phi}XH^2(3G_{5X} + 2XG_{5XX}), \tag{1.119}$$

$$M_*^2\alpha_T = 2X(2G_{4X} - 2G_{5\phi} - (\ddot{\phi} - \dot{\phi}H)G_{5X}). \tag{1.120}$$

The ΛCDM model is given by $\forall i.\alpha_i = 0$. (Note that vanishing α-functions does not mean that DE disappears! It merely reduces it to a cosmological constant with $\Omega_\Lambda = \Omega_\varphi$.) Although we have just four functions of time to consider, as free functions they represent infinite degrees of freedom. This is not a space which we can meaningfully sample. In order to obtain any results, a parameterisation $\alpha_i(t) = c_iF(t)$ is required. One such parameterisation, used frequently in the literature, is $\alpha_i(t) = c_i\Omega_\varphi(t)$ [132][137]. Now we need only sample over four additional parameters, c_i, but we must bear in mind that this is a very reduced part of the overall theory space. Nevertheless, it allows us to check for deviations from GR in a way that bears some physical interpretation. Each of the functions α_i has a physical interpretation based on its origin and effect in the field theory. As such, if we find that one of these parameters is significantly non-zero, then we know something about the direction in which to build our new gravitational theories!

Despite the fact that the starting point of this approach – the Horndeski class – encompasses many popular theories, the need for a realistic course of action forces us towards a phenomenological approach. This reduced subclass of theories, involving $a_i(t) \propto \Omega_\varphi(t)$, does not directly contain theories such as $f(R)$ or quintessence. Although specific theories will not be modelled in this thesis, it is interesting to consider how they map into the α-function formalism, and how closely they might resemble a chosen time dependence.

1.2.3.1 Horndeski in the wake of gravitational waves

A recent detection of a gravitational wave (GW) event with an electromagnetic counterpart [150] has placed strong constraints on the deviation of gravitational wave speed from the speed of light (at least at the present) [151]. This has been converted to a constraint on $\alpha_T \lesssim 10^{-15}$ at the present epoch [148][149]; since the models we consider are monotonic increasing this implies even stronger constraints in the past. As such, one can safely set $\alpha_T = 0$ in any numerical cosmological analysis, since there is no hope of such a tiny deviation rising above numerical or observational noise. What does this mean for modified gravity?

In terms of a parameterised analysis, it simply knocks out one of the sampling variables. If we want to understand that this means for our action, we can look at α_T,

$$\alpha_T = 2M^{-2}X(2\frac{\partial G_4}{\partial X} - 2\frac{\partial G_5}{\partial \phi} - (\ddot{\phi} - \dot{\phi}H)\frac{\partial G_5}{\partial X}) = 0. \tag{1.121}$$

The simplest way to eliminate α_T is to have $\partial_X G_4$, $\partial_\phi G_5$, and $\partial_X G_5$ all vanish identically. This means that in our action, we may only couple to R using a function of ϕ alone (with no kinetic term, and we can have at most a constant coupling to the Einstein tensor $G_{\mu\nu}$. This greatly simplifies the possible actions we may explore, although it is not such a great blow to the theories present in the literature. Whilst some models (such a quartic and quintic Galileons) are disposed of, most theories stick to a more conservative Lagrangian that meets these conditions, and as such symmetrons, dilatons, and $f(R)$ (to name but a few) all survive intact.

We may even take a less severe approach, and ask only that the terms con-

tributing to α_T cancel algebraically at all times (or, at least, at very recent times). This approach is not terribly complicated if $\partial_X G_5 \equiv 0$. Then we only require that $\partial_X G_4 = \partial_\phi G_5$, which is easy to construct and would trivially hold at all times. (Things are much more difficult if we allow G_5 to be a function of X, as then we must have a delicate cancelation which also includes ϕ, its derivatives, and H at all times. We will not pursue this further as the actions which we consider explicitly in Part II contain no such couplings, and the analysis we perform in Part I may be achieved by setting $\alpha_T = 0$ numerically.

1.3 Structure and Dynamics in the Local Universe

The local universe is somewhat on the 'edge' of cosmology, being the smallest scales that we can look at and still attempt to call ourselves cosmologists. It stretches out on the scales of a megaparsec up to many hundreds of megaparsecs, and covers the structures in our immediate vicinity. Galaxies and their satellites, galaxy clusters, filamentary structures, all the way up to superclusters may be included on these scales. The interaction between this scale and the traditional idea of cosmology – i.e. the universe as a whole – is very interesting [99]. What can we learn about the universe from the local volume, which we can observe in the greatest detail? How does our understanding of cosmology help us to understand formation and behaviour of the local volume? We will find that even on scales of just a megaparsec, absurdly small on 'cosmological scales', our understanding is inextricably linked to our fundamental models of cosmology.

Unlike the large scale universe, the local universe is highly inhomogeneous and anisotropic. These structures that we see in our local universe are fascinating but pose great theoretical challenges, being formed in highly non-linear processes and without simplifying symmetries. For these reasons we must often rely on n-body simulations for physics on these scales which cannot be treated analytically or perturbatively.

1.3.1 Motion of Matter in Static and Co-Moving Coordinates

Matter in weak fields may be approximated by Newtonian dynamics, provided the scales considered are much larger than the Schwarzschild radii of any collapsed objects, and small enough that the Hubble velocities are much less than the speed of light (so we are considering 'non-relativistic' speeds). This actually applies to the bulk of structure formation theory and n-body simulations. Although we are treating the problem in a Newtonian way, and in a flat, Euclidean space, we will see that co-moving coordinates are still a very useful tool for understanding the evolution of perturbations against a background of expansion.

We have already seen that the weak field limit of GR is a modified Poisson equation

$$\nabla_r^2 \Phi = 4\pi G(\rho + 3pc^{-2}) - \Lambda, \tag{1.122}$$

which reduces to Newtonian gravity in the limit $\Lambda \to 0$, $p \to 0$. For an spherical distribution of pressureless matter, this results in the modified acceleration equation

$$\ddot{r} = -\frac{GM_r}{r^2} + \frac{\Lambda c^2}{3} r. \tag{1.123}$$

We shall keep the Λ terms throughout, as these are not negligible on e.g. intergalactic or galaxy clustering scales.

We may also consider dynamics of a particle on a uniform background, following Peebles (1971) [56]. Given

$$\nabla_r^2 \Phi = 4\pi G\rho_b - \Lambda, \tag{1.124}$$

for uniform ρ_b, one may integrate to find that

$$\Phi = \frac{2\pi}{3} G\rho_b r^2 - \frac{\Lambda}{6} r^2. \tag{1.125}$$

Using now that $\ddot{r} = -\frac{\partial \Phi}{\partial r}$,

$$\ddot{r} = \left[-\frac{4\pi}{3} G\rho_b + \frac{\Lambda}{3} \right] r, \tag{1.126}$$

which is just an equation for uniform distance scaling ($\ddot{r} \propto r$). We may transform to comoving coordinates $r = a(t)x$, where x is constant in which case the equation dictates the scale factor $a(t)$,

$$\ddot{a} = \left[-\frac{4\pi}{3} G\rho_b + \frac{\Lambda}{3} \right] a. \tag{1.127}$$

Now considering the peculiar motion of a particle with position $\vec{x}$, we have

$$\mathscr{L} = \frac{1}{2} m\dot{r}^2 - m\Phi = \frac{1}{2} m(\dot{a}x + a\dot{x})^2 - m\Phi. \tag{1.128}$$

This Lagrangian can be made more palatable by subtracting a total time derivative (which must leave the equations of motion unchanged)

$$\mathscr{L}' = \mathscr{L} - \frac{d}{dt}\left(\frac{1}{2} ma\dot{a}x^2 \right), \tag{1.129}$$

$$= \frac{1}{2} ma^2\dot{x}^2 - \frac{1}{2} ma\ddot{a}x^2 - m\Phi, \tag{1.130}$$

$$= \frac{1}{2} ma^2\dot{x}^2 - m\Phi', \tag{1.131}$$

where $\Phi' = \Phi + \frac{1}{2} a\ddot{a}x^2$ defines a new effective potential. We may formulate the Poisson equation for this new potential in comoving coordinates by taking $\nabla_x^2 \Phi'$,

$$\nabla_x^2 \Phi' = a^2 \nabla_r^2 \Phi + \nabla_x^2 \left(\frac{1}{2} a\ddot{a}x^2 \right), \tag{1.132}$$

$$= 4\pi G\rho a^2 - \Lambda a^2 + 3a\ddot{a}. \tag{1.133}$$

We may now substitute in the scale factor acceleration equation,

$$\nabla_x^2 \Phi' = 4\pi G a^2 (\rho - \rho_b), \tag{1.134}$$

and we see that the Λ terms cancel out entirely in comoving coordinates, since the distance scaling from the cosmological constant affects the background in exactly the same way as a test particle (and, indeed, all particles are treated the same by Λ). As a result, in cosmological simulations where motions are expressed in comoving

coordinates, we never see the Λ contribution to the acceleration equation explicitly, but it has been absorbed into the scale factor $a(t)$.

1.3.2 The Local Group

The Local Group (LG) is the collection of galaxies in our neighbourhood, including our own. By far its most massive components are the Milky Way (MW, our home) and our nearest fully fledged neighbour, Andromeda (M31). There are a number of other small galaxies and dwarf satellites which make of the rest of the mass, of which the Triangulum galaxy (M33), and the Large and Small Magellanic Clouds are probably the most note-worthy. As of 2003, there were 36 identified systems in the LG [110]; recently DES has spotted a number of additional dwarf satellites [111]. Despite being the region of the universe which is closest to us, it remains fairly mysterious. The galactic plane obscures our view of the sky, and unlike in large scale cosmology, we cannot make assumptions about what is there based on statistics! In the LG, we are far from the homogeneous models that we inhabit in cosmology.

Even the most fundamental questions about the LG galaxies – what are their masses? what are their relative motions? – remain difficult to answer with great certainty, and there are many conflicting results in the literature ([98][118][120][81] and more). One might ask (particularly if one considers oneself a cosmologist in the purest sense of the word), why on Earth anyone should care to investigate the LG. There are, after all, many galaxy clusters which will have slightly different properties due to this and that random happenstance and the configuration that our home has fallen into is unlikely to tell us much about fundamental physics. Well, there may be something to that, but nevertheless: it's my home, and I want to know what's going on.

This is not to say that it is not *connected* with cosmology. We shall see that our theory of gravity, and the development of the large scale structure in which the LG is embedded, have significant effects on our understanding of the LG, and our estimations of its properties.

1.3.3 The Cosmic Web & Cosmography

The 'cosmic web' (CW) is the name given to the observed structure in the universe on a scale below homogeneity, but larger than galaxy clusters. When we look at the patterns of matter density in the universe – via, for instance, galaxy positions – or in numerical simulations, we see filamentary structure; there are strands of matter stretching through space, meeting at intersections ('knots'), and leaving holes between them ('voids'). Matter may also move down these filaments towards the knots, as is the case of the LG moving towards the Virgo cluster.

The definition above is certainly lacking in rigour, and this is a characteristic problem in study of the CW. Classification of the CW into knots, filaments, walls, and voids is a significant challenge, particularly since its definition is based on simply eyeballing the data. The best we can do is find a robust way of characterising the matter distribution which matches our visual interpretation. Different methodologies exist, which may produce different results. Some are discrete, and involve creating a graph (a set of nodes and edges between them), which allows one to study topological aspects of the web (such as the degree of nodes), but doesn't allow us to identify some of the structure (such as walls) or understand its continuous physical properties [103]. One continuous method is to analyse a tensor defined over the space, and use the eigenvectors and eigenvalues to characterise the structures [105][104][106]. Their properties can allow the distinction of structural categories, but the continuity allows one to study the properties of said structures and the way they change across space as well. The tensors which can be successfully used are the tidal tensor (T_{ij}, characterising the gravitational field) and the velocity shear tensor (Σ_{ij}, characterising the velocity flows of galaxies), which are not unrelated.

Given a set of eigenvectors $\vec{e}_i$ and corresponding eigenvalues λ_i, and some threshold value λ_*, we can describe the structure as follows:

- $\lambda_{0,1,2} \leq \lambda_*$ characterises a void,
- $\lambda_0 > \lambda_*$, $\lambda_{1,2} \leq \lambda_*$ characterises a wall,
- $\lambda_{0,1} > \lambda_*$, $\lambda_2 \leq \lambda_*$ characterises a filament,

- $\lambda_{0,1,2} > \lambda_*$ characterises a knot,

where $\lambda_0 \geq \lambda_1 \geq \lambda_2$. The sign of the eigenvalue describes tendency for matter to either move together or apart along the direction of that eigenvector: when matter is moving apart in all directions we have a void, and when it is moving together in all directions it is a knot. A filament shows compression along two axes, creating its cross section, and expansion along the either, which defines line of the filament. The magnitude of the eigenvalue demonstrates the strength of the collapse or expansion, which is not available in network methods.

Cosmography is the mapping of the universe, particularly the local universe. It involves not only placing the galaxies in space, but also understanding their motions. This information can then be used to better understand our own extragalactic system, as well as make cosmological inferences. Information about galaxy flows can, in particular, be very useful in measuring local values of H_0, which has a notorious tension with the comparatively low CMB-based estimates from Planck. Such systems can also be used to study other aspects of standard cosmology, and may even be useful for studying such exotic phenomena as DE and MG.

The picture of the local universe starts with our own galaxy, which sits amongst our close neighbours (the LG), which sits inside a filament in the cosmic web, flowing towards the Virgo supercluster (swept up in the 'Virgocentric flow'), and larger even than that the Laniakea supercluster [107].

1.4 Large Scale Structure

The primary source of cosmological information considered in this thesis will be the formation of large scale structure (LSS) in the universe. Looking out to redshift $z \approx 2$ allows us to see over a substantial fraction of the universe's history (some 10 Gyr or so depending on cosmology) and thus allows us to track, on a statistical level, gravitational collapse on different scales and over time. Since most matter is dark, the matter distribution must be inferred by looking to galaxies; on the one hand we may look at their distribution across the sky and in redshift $n(\Omega, z)$, and on the other we may look for distortions in their ellipticity as the light they emit is

lensed by the matter along the line of sight.

1.4.1 Structure Formation Theory Basics

Large scale structure (LSS) in the universe is seeded by quantum fluctuations in the very early universe [26][29]. For example, fluctuations in a scalar inflation field $\delta\phi(t,\vec{x})$ will result in slightly different local expansion rates. Starting from a homogenous matter distribution, this results slightly different local densities; overdense regions will then gravitate in excess of the background and attract more matter towards them. The amplitude of scalar curvature fluctuations is not set by the standard model of cosmology, but is a free parameter A_s, sometimes denoted Δ_R^2. The k-dependence of the fluctuation amplitude is typically taken to be a power law, where [56]

$$\Delta_R^2(k) = \Delta_R^2(k_*)\left[\frac{k}{k_*}\right]^{n_s-1}. \tag{1.135}$$

The amplitude of fluctuations in the model is fixed at some scale k_*, and then the amplitude at other scales is derived from the power law. (There is, unfortunately, no set standard for k_*, so comparisons of results must take this into account.) The spectral index n_s is close to unity, and for a scale independent spectrum $n_s = 1$.

Once the primordial power spectrum is calculated, the growth of perturbations can be tracked over universal time, until reaching the present. The linear growth, which accounts for larger scale structure (but requires non linear correction at smaller scales), is governed by the perturbation equations [56][24]

$$\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} = \frac{1}{\bar{\rho}a^2}\nabla^2 p + 4\pi G\bar{\rho}\delta, \tag{1.136}$$

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\vec{v} = 0. \tag{1.137}$$

The solution to this equation gives the linear growth function $D(z)$, where we parameterise in terms of z rather than t as this is for most purposes in both theory and data analysis more convenient. This evolving matter distribution will determine the placement, and the lensing, of the galaxies which we observe. This, coupled with information from the CMB, will allow us to constrain our models of cosmology.

1.4.2 Correlation Functions and the Power Spectrum

The structure which has developed must be analysed statistically – our hypotheses about the early universe are effectively statistical, involving global properties and stochastic processes. Cosmological theory does not really cover the formation of compact objects such as galaxies, which is a field in itself with connections to cosmology which are still poorly understood. This is particularly true when we consider that the bulk of matter in the universe is dark, and thus the structures that we observe make up only a small part of the universe. A statistical approach must therefore be taken to observables, and in cosmology we are often interested in correlation functions as a measure of structure. We instead assume that astrophysical objects trace the underlying matter density distribution, in other words [56]

$$P\delta V \propto \delta(\mathbf{x}), \tag{1.138}$$

where $P\delta V$ is the probability of finding an object in an infinitesimal volume δV, and $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}}$ is the matter overdensity field. In order to make predictions, we must work with statistics regarding the overall matter distribution, and then observable statistics (e.g. galaxy correlations) are inferred form that. The correlation function $\xi(r)$ of a continuous field is defined as [25]

$$\xi(\vec{r}) = \langle \delta(\vec{x})\delta(\vec{x}+\vec{r}) \rangle. \tag{1.139}$$

Since one can expand $\delta(\vec{r})$ in terms of its Fourier components:

$$\xi(\vec{r}) = \langle \frac{1}{(2\pi)^3} \int \delta_{\vec{k}} \delta^*_{\vec{k}} e^{-i\vec{k}\cdot\vec{r}} d^3k \rangle \tag{1.140}$$

we see that the correlation function is the Fourier transform of the power spectrum (and vice versa), where the power spectrum is:

$$P(k) = \langle |\delta_k|^2 \rangle = FT\left[\xi(\vec{r})\right] \tag{1.141}$$

1.4.2.1 The Angular Power Spectrum $C(l)$

The angular power spectrum, when decomposed into spherical harmonics, is denoted $C(l)$. The $C(l)$ formalism proves to be very useful in analysing cosmological data. For any two probes α and β (say galaxy number counts, weak lensing shear, etc.) and two redshift bins i and j, then the angular power spectrum $C^{ij}_{\alpha\beta}(l)$ can be calculated [46],

$$C^{ij}_{\alpha\beta}(l) = \frac{2}{\pi}\int W^i_\alpha(k,l)W^j_\beta(k,l)P(k)k^2\,dk\,. \tag{1.142}$$

This is clearly symmetric in $(i \leftrightarrow j)$, and in $(\alpha \leftrightarrow \beta)$. This formalism makes it straightforward to calculate auto- and cross-correlations between redshift bins and cosmological probes. The form of the window functions $W^i_\alpha(k,l)$ will be given in later chapters.

The $C(l)$ may be calculated as follows (see e.g. [25]). The observed overdensity in a direction $\hat{n}$ is

$$\delta(\hat{n}) = \int_0^\infty \delta(\vec{\chi})\phi(\chi)d\chi, \tag{1.143}$$

where $\phi(\chi)$ is the radial selection function for the sample, $\vec{\chi}$ is the comoving distance vector (with the observer location at $\vec{\chi} = \vec{0}$), and $\chi = |\vec{\chi}|$. (Here we have implicitly included the factor of χ^2 from the volume integral into the selection function.) We may also expand the matter perturbation in Fourier space,

$$\delta(\vec{\chi}) = \int \delta(\vec{k})e^{i\vec{k}\cdot\vec{\chi}}\frac{d^3k}{(2\pi^3)} = 4\pi\int \delta(\vec{k})i^l j_l(k\chi)Y_{l,m}(\hat{\chi})Y^*_{l,m}(\hat{k})\frac{d^3k}{(2\pi^3)}, \tag{1.144}$$

where $\hat{\chi}$ and $\hat{k}$ are unit vectors in the direction of $\vec{\chi}$ and $\vec{k}$, and we have used the expansion of a plane wave in spherical harmonics and Bessel functions. Expanding

also $\delta(\vec{\chi})$ in spherical harmonics, we have by orthogonality:

$$a_{l,m} = \int Y^*_{l,m}(\hat{n})\delta(\hat{n})d\Omega, \tag{1.145}$$

$$= \int Y^*_{l,m}(\hat{n}) \left[\int \delta(\vec{\chi})\phi(\chi)d\chi \right] d\Omega, \tag{1.146}$$

$$= 4\pi \int Y^*_{l,m}(\hat{n}) \left[\int \left(\int \delta(\vec{k}) i^{l'} j_{l'}(k\chi) Y_{l',m'}(\hat{\chi}) Y^*_{l',m'}(\hat{k}) \frac{d^3k}{(2\pi^3)} \right) \phi(\chi)d\chi \right] d\Omega, \tag{1.147}$$

$$= 4\pi \int \phi(\chi) \int \delta(\vec{k}) i^l j_l(k\chi) Y^*_{l,m}(\hat{k}) \frac{d^3k}{(2\pi^3)}\, d\chi\,. \tag{1.148}$$

In the last step we have integrated over the solid angle and used the orthogonality of spherical harmonics. We shall now take the expectation value of $a_{l,m}$ with $b^*_{l,m}$, where $b_{l,m}$ are the spherical components of a probe sourced from the same matter distribution, but which may have a different selection function $\psi(\chi)$ i.e. it may be either an auto or cross correlation.

$$\langle a_{lm}\, b^*_{lm} \rangle = \left\langle 16\pi^2 \int \delta(\vec{k})\, Y^*_{lm}(\vec{k}) \int \phi(\chi) j_l(k\chi) d\chi \frac{d^3k}{(2\pi^3)} \right. \\ \left. \times \int \delta^*(\vec{k'}) Y_{lm}(\vec{k'}) \int \psi(\chi') j_l(k'\chi') d\chi' \frac{d^3k'}{(2\pi^3)} \right\rangle. \tag{1.149}$$

The only stochastic quantity here is δ, so using the linearity of the expectation operator we may take it inside the integrals and furthermore appeal to the fact that $\delta(k)$ has a homogenous random distribution and thus

$$\langle \delta(k)\delta^*(k') \rangle = \begin{cases} P(k)(2\pi)^3 & k = k', \\ 0 & k \neq k'. \end{cases} \tag{1.150}$$

(This may be expressed succinctly using a Dirac delta function, but given the notation for the perturbation has been avoided.) Thus integrating over k' yields the

simpler expression:

$$\langle a_{lm}\, b^*_{lm}\rangle = 16\pi^2 \int P(k)|Y_{lm}(\vec{k})|^2 \int \phi(\chi) j_l(k\chi) d\chi \int \psi(\chi') j_l(k\chi') d\chi' \frac{d^3k}{(2\pi^3)}. \tag{1.151}$$

Note that this makes use of the isotropy of $P(k)$, and that although we have now only one k variable, we still have separate χ and χ' for the window functions. Summing over m and using the spherical harmonic summation

$$\sum_{m=-l}^{l} |Y_{lm}(\vec{n})|^2 = \frac{2l+1}{4\pi}, \tag{1.152}$$

$$C_l = 4\pi \int P(k) \int \phi(\chi) j_l(k\chi) d\chi \int \psi(\chi') j_l(k\chi') d\chi' \frac{d^3k}{(2\pi^3)}. \tag{1.153}$$

Expanding the k-volume integral and using spherical symmetry, we have

$$C_l = \frac{2}{\pi} \int P(k) \left[\int \phi(\chi) j_l(k\chi) d\chi\right] \left[\int \psi(\chi') j_l(k\chi') d\chi'\right] k^2 dk. \tag{1.154}$$

We see that our window functions are integrals over our radial selection functions, modulated by spherical Bessel functions.

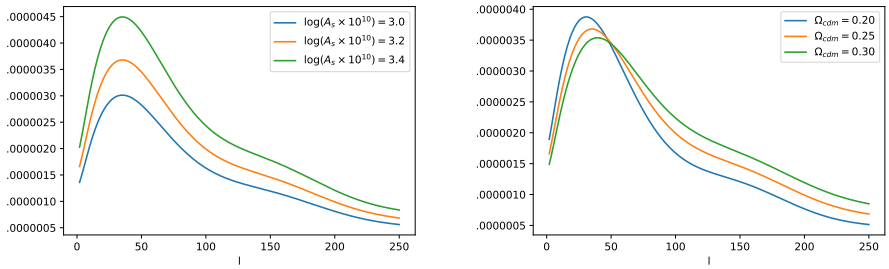
1.4.2.2 Dependencies on Cosmology

These power spectra contain a wealth of information about cosmology and fundamental physics, from its matter-energy composition to the nature of gravity. The cosmologies that we consider will be parameterised by the following variables:

- A_s; the amplitude of initial fluctuations, this determines the height of the primordial power spectrum. It can be replaced by the normalisation σ_8 if preferred.

- Ω_i; the fractional density of component i of the universe's energy. This includes b for baryons, cdm for cold dark matter, Λ for the cosmological constant. This can be extended to include r (radiation), ν (neutrinos), and ϕ (scalar fields).

- n_s; the scalar spectral index. The amplitude of primordial fluctuations



(a) **(b)**

Figure 1.1: Variation of an autocorrelation $C(l)$ of galaxy number counts with the height of the power spectrum (A_s) and the dark matter content of the universe Ω_{cdm}. A_s changes the height but not the shape of the $C(l)$ function, because it only influences the initial amplitude. Ω_{cdm} on the other hand is involved with how the spectrum evolved, and can change the shape considerably.

$$\propto \left(\frac{k}{k^*}\right)^{n_s}.$$

- h; equivalent to $H_0/100$ (H_0 in units of km s^{-1} Mpc^{-1}).

- τ_r; the optical depth at reionisation.

The sensitivity of the $C(l)$s to these parameters will depend on both how they affect the primordial spectrum $P(k)$, and how they affect the growth of structure. As an example, fig 1.1 shows how the galaxy number counts $C(l)$ is affected by the cosmological parameters A_s and $\Omega_{\rm cdm}$. A_s controls the amplitude of the power spectrum, and thus also controls the amplitude of the $C(l)$; $\Omega_{\rm cdm}$ affects the height and the shape of the $C(l)$ since it changes both the structure formation and the expansion history of the universe through the balance of Ω_m and Ω_Λ. The more similar the effects on the $C(l)$s are from two different parameters, the more those parameters will appear degenerate in the analysis. For this reason, it is useful to combine multiple probes of cosmology, as parameters which are degenerate in one observable may not be in another.

As we have seen from equation 1.154, the calculation of theoretical $C(l)$ values requires knowledge of the redshift distribution of galaxies which we will be analysing, as the formation of structure will have progressed to different extents at different times (redshifts). Therefore, in order to compare the data and the theory,

we must have reliable redshift information, at least at the level of the probability distribution (which, thankfully, is a much weaker condition than having reliable information for individual objects). For spectroscopic samples, we have good quality information about each object, and therefore can reconstruct any probability distributions with great accuracy. The same is not true of photometric redshifts (which we will discuss a little later), which have associated with them large random and systematic errors. We shall explore the calibration of these redshifts in the chapters which follow.

1.4.3 Weak Gravitational Lensing

Weak gravitational lensing is a valuable cosmological observable which has been gaining in importance in the field year after year. Observing weak lensing is exceedingly difficult due to the intrinsic shapes and alignments of galaxies, and noisy imaging data. Nevertheless, the payoff is significant. In GR, weak lensing simply traces the matter distribution along the line of sight, but unlike galaxy number counts there is no highly uncertain bias term to worry about. When combined with position information it is possible to constrain the little understood galaxy bias which can lead to spurious inference if not properly modelled. Lensing becomes even more useful in the context of modified gravity; we shall see below that lensing can detect 'anisotropic stress' in MG/DE models. (Anisotropic stress can also occur from the matter sector if one proposes some unconventional energy-momentum tensors.) The derivation presented here will follow [31] and [34]. Overviews can also be found in [29][25]. We shall work in a metric

$$ds^2 = (1+2\Psi)dt^2 - (1-2\Phi)d\vec{x}^2, \tag{1.155}$$

which is a general perturbation to a flat metric in the Newtonian gauge. In GR in the absence of anisotropic stress $\Psi = \Phi = 4\pi G\rho$. We shall retain the distinct 'Bardeen potentials' for the sake of applicability to more general theories. It will also be worth noting ahead of time that, in a MG gravity theory, we may have a modified

relativistic Poisson equation

$$\nabla^2(\Psi+\Phi) = \Sigma(z) 8\pi G \rho_m. \tag{1.156}$$

In GR, the modification term $\Sigma(z)$ is unity. This is a general parameterisation which is frequently used; in principle Σ is an arbitrary function which can, in principle, be derived for a particular theory.

Weak gravitational lensing is concerned with the bending of light sourced by the presence density fluctuations, due to perturbations to the metric. Light travels along null geodesics, and therefore we have

$$(1+2\Psi)c^2dt^2 = (1-2\Phi)d\vec{x}^2 \tag{1.157}$$

$$\implies c^2dt^2 = (1-2\Phi)(1+2\Psi)^{-1}d\vec{x}^2. \tag{1.158}$$

In weak lensing we assume that the metric perturbations are smalls $\Phi \ll 1$, $\Psi \ll 1$. We can then write to first order

$$c^2\,dt^2 = (1-2(\Phi+\Psi))\,d\vec{x}^2, \tag{1.159}$$

$$c\,dt = (1-(\Phi+\Psi))\,d\vec{x}. \tag{1.160}$$

Following [34] the lensing effect can be derived using Fermat's principle of least time,

$$t = \frac{1}{c}\int (1-[\Phi+\Psi])\,d\vec{x}. \tag{1.161}$$

This means that the effective refractive index of the geometry is

$$n(\vec{x}) = (1-[\Phi+\Psi]). \tag{1.162}$$

Progress is most easily made if the curve is parameterised with respect to some

parameter σ, such that

$$d\vec{x} = \left|\frac{d\vec{x}}{d\sigma}\right| d\sigma, \tag{1.163}$$

$$t = \frac{1}{c}\int n\frac{d\vec{x}}{d\sigma}, \tag{1.164}$$

$$\mathscr{L} = n|\dot{\vec{x}}|. \tag{1.165}$$

The Euler-Lagrange equations state

$$\frac{d}{d\sigma}\left(\frac{\partial\mathscr{L}}{\partial\dot{\vec{x}}}\right) - \frac{\partial\mathscr{L}}{\partial\vec{x}} = 0. \tag{1.166}$$

We know that $\frac{\partial\mathscr{L}}{\partial\vec{x}} = \vec{\nabla}(1-[\Phi+\Psi])|\dot{\vec{x}}|$, and that $\frac{\partial\mathscr{L}}{\partial\dot{\vec{x}}} = (1-[\Phi+\Psi])\frac{\dot{\vec{x}}}{|\dot{\vec{x}}|}$, and by definition $\dot{\vec{x}}$ is tangent to the curve. We can always choose a σ such that $|\dot{\vec{x}}| = 1$ so that $\dot{\vec{x}}$ is the unit tangent to the curve. Putting all of this together we have

$$\frac{d}{d\sigma}(n\vec{e}) - \vec{\nabla}n = 0. \tag{1.167}$$

We may then say that

$$\frac{\partial n}{\partial\sigma}\vec{e} + n\frac{\partial\vec{e}}{\partial\sigma} = \vec{\nabla}n. \tag{1.168}$$

Using the fact that

$$\frac{\partial n}{\partial\sigma} = \frac{\partial n}{\partial\vec{x}}\frac{\partial\vec{x}}{\partial\sigma} = \vec{\nabla}n\cdot\vec{e}, \tag{1.169}$$

then we have

$$n\dot{\vec{e}} + (\vec{\nabla}n\cdot\vec{e})\vec{e} = \vec{\nabla}n. \tag{1.170}$$

The term $(\vec{\nabla}n\cdot\vec{e})\vec{e}$ has an important physical significance. $(\vec{\nabla}n\cdot\vec{e})$ is the component of the derivative of n in direction of the tangent vector, so $(\vec{\nabla}n\cdot\vec{e})\vec{e}$ is $\vec{\nabla}_{\|}n$ the derivative vector of n in the direction of the path. We can write $\vec{\nabla}n - \vec{\nabla}_{\|} = \vec{\nabla}_{\perp}$ and thus

$$n\dot{\vec{e}} = \vec{\nabla}_{\perp}n \implies \dot{\vec{e}} = \vec{\nabla}_{\perp}\ln(n), \tag{1.171}$$

and using $\ln(n) \approx -(\Phi + \Psi)$ then

$$\hat{\vec{\alpha}} = -\int_{\sigma_A}^{\sigma_B} \vec{\nabla}_\perp (\Phi + \Psi)\, d\sigma, \tag{1.172}$$

i.e. the change in angle of the light ray between emission and reception is the integral of the changes in direction along the path of the light, assuming the small angle approximation.

The co-moving displacement is $d\vec{X}(\chi) = f_K(\chi - \chi')d\hat{\alpha}$ using the small angle approximation. (The displacement $\vec{X}$ is 2D, as is α, and is not the same as the comoving position vector $\vec{x}$.) The difference between two light rays, one traversing a path with potentials Φ^0 and Ψ^0 is

$$\vec{X}(\chi) = f_K(\chi)\vec{\theta} - \int_0^\chi f_K(\chi - \chi')(\nabla_\perp[\Phi + \Psi] - \nabla_\perp[\Phi^0 + \Psi^0])\, d\chi'. \tag{1.173}$$

In a case with no lensing, there would be an angle $\vec{\beta} = \frac{\vec{x}}{f_K(\chi)}$; the difference between the observed angle and the unlensed angle is $\vec{\alpha} = \vec{\theta} - \vec{\beta}$.

$$\vec{\alpha} = \int_0^\chi \frac{f_K(\chi - \chi')}{f_K(\chi)} (\nabla_\perp[\Phi + \Psi] - \nabla_\perp[\Phi^0 + \Psi^0])\, d\chi'. \tag{1.174}$$

The expression above is very tricky to integrate, and so we apply the Born approximation assuming small perturbations. The effects of the lensing on a 2D separation vector can be placed in a matrix

$$A_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j}. \tag{1.175}$$

In the absence of lensing $\vec{\alpha} \to \vec{0}$ and $\mathbf{A} \to \mathbb{I}$. This is the Jacobean matrix for a 'coordinate transform' from $\vec{\theta}$ to $\vec{\beta}$ (although we are not strictly performing a coordinate transform it can be cast in this way). Using the Born approximation we may say that $\vec{x} \approx f_K(\chi)\vec{\theta}$ i.e. we approximate the separation as an unperturbed ray. Then

$\frac{\partial}{\partial\theta_i} = f_K(\chi)\frac{\partial}{\partial x_i}$, and we may calculate the amplification matrix A_{ij} in full,

$$A_{ij} = \delta_{ij} - \int_0^\chi \frac{f_K(\chi-\chi')}{f_K(\chi)} f_K(\chi') \left(\frac{\partial^2}{\partial x_i \partial x_j}[\Phi+\Psi]\right) d\chi' . \tag{1.176}$$

The second term in this equation can be written as the 2D Laplacian of the so-called 'lensing potential' ψ,

$$A_{ij} = \delta_{ij} + \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}, \tag{1.177}$$

$$\psi = \int_0^\chi \frac{f_K(\chi-\chi')}{f_K(\chi)f_K(\chi')}[\Phi+\Psi]\, d\chi' . \tag{1.178}$$

The matrix A_{ij} is clearly symmetric, and thus can be split into a diagonal and trace-free part. The diagonal part is known as the *convergence*, which acts as a magnification, and the trace free part is the *shear*, which distorts the image,

$$\mathbf{A} = \mathbb{I} + \begin{pmatrix} -\kappa & 0 \\ 0 & -\kappa \end{pmatrix} + \begin{pmatrix} -\gamma_1 & -\gamma_2 \\ -\gamma_2 & \gamma_1 \end{pmatrix}. \tag{1.179}$$

The convergence and the shear are not independent, and we shall focus on the measurement of the shear matrix. The convergence matrix can clearly be represented as a single real scalar κ, as it contains no directional information. The shear matrix may be described as a 'spin-2 field' (although it has nothing to do with the quantum notion of spin). The shear matrix is a real, symmetric matrix, with eigenvalues $\lambda_{1,2} = \pm\sqrt{\gamma_1^2+\gamma_2^2}$. This means it is diagonalisable by a real orthogonal matrix, and may be represented as

$$\gamma = \lambda \mathbf{R}^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{R}. \tag{1.180}$$

where $\mathbf{R}$ is the 2D rotation matrix $\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$. This results in a shear matrix which can be written as

$$\gamma = \lambda \begin{pmatrix} \cos 2\phi & -\sin 2\phi \\ -\sin 2\phi & -\cos 2\phi \end{pmatrix}. \tag{1.181}$$

Therefore we see that all the information in the shear field may be captured in two values, a magnitude and an orientation, which is invariant under rotations of π – hence spin-2. We may then write the shear as $\lambda e^{2i\phi}$, which captures the information and the transformation properties of the shear field. We can also calculate the angle ϕ of the rotation matrix and find

$$\tan 2\phi = \frac{\gamma_2}{\gamma_1}, \tag{1.182}$$

and hence the shear can also be written as a complex number $\gamma_1 + i\gamma_2$ with a polar angle of 2ϕ.

The convergence κ is clearly related to the 2D Laplacian of the lensing potential $\kappa = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\theta_1^2} + \frac{\partial^2\psi}{\partial\theta_2^2}\right)$. We would like to convert the 2D Laplacian of ψ to the 3D Laplacian of $[\Phi+\Psi]$ in order to get the lensing potential in terms of the matter distribution, rather than the metric potentials. This is acceptable provided that the additional $\frac{\partial^2}{\partial\chi^2}$ term is small integrated along the line of sight. The integral $\int \frac{\partial^2[\Phi+\Psi]}{\partial\chi^2}\,d\chi \sim \left[\frac{\partial[\Phi+\Psi]}{\partial\chi}\right]_{\chi_1}^{\chi_2}$, and for a bound object $\frac{\partial[\Phi+\Psi]}{\partial\chi} \approx 0$ outside as the space returns to the background solution. So this part of the integral will be approximately zero if the extent of the object is small compared to the length scale of variations in the geometric term $\frac{f_K(\chi-\chi')}{f_K(\chi)f_K(\chi')}$.

Then we have an expression for the convergence for sources in direction $\vec{\theta}$ at a comoving distance χ,

$$\kappa(\vec{\theta},\chi) = \int_0^\chi \frac{f_K(\chi-\chi')}{f_K(\chi)} f_K(\chi')\Sigma(\chi')\delta(\vec{\theta},\chi')\,d\chi'. \tag{1.183}$$

This can be integrated over $n(\chi)\,d\chi$ to get the average convergence $\kappa(\vec{\theta})$ for a tomographic bin. Notice that there is an additional $\Sigma(\chi)$ term compared to the standard GR lensing equation. Lensing is therefore sensitive to modified growth via δ, and sensitive to gravitational slip via Σ; the galaxy number counts is sensitive only to the modified growth. This has given us a lensing expression in terms of the matter distribution, but we are more interested in the shear field. Fortunately, the power spectrum of the shear field is the same as the power spectrum of the

convergence field. (The $C(l)$s are related by a simple combination of l-factors.) In Fourier space, we may express the convergence and shear in terms of the conjugate variable of $\vec{\theta}$, the 2D wavevector $\vec{y}$. Then the expressions for shear and convergence are

$$\tilde{\kappa}(\vec{y}) = \frac{1}{2}|\vec{y}|^2\psi, \tag{1.184}$$

$$\tilde{\gamma}_1(\vec{y}) = \frac{1}{2}(y_1^2 - y_2^2)\psi, \tag{1.185}$$

$$\tilde{\gamma}_2(\vec{y}) = y_1 y_2 \psi. \tag{1.186}$$

Using the complex number representation of the shear field, $\gamma = \gamma_1 + i\gamma_2$, we find that

$$\gamma = \frac{1}{2}(y_1 + iy_2)^2\psi, \tag{1.187}$$

and therefore

$$\gamma = \frac{(y_1 + iy_2)^2}{y_1^2 + y_2^2}\kappa = \kappa e^{2i\varphi}, \tag{1.188}$$

where we have written $(y_1 + iy_2)$ in polar form as $|y|e^{i\varphi}$. This means that the convergence and the shear are related by a complex phase, and the power spectrum for γ and κ are the same.

Lensing can, in GR, be used to constrain the galaxy bias $\delta_g = b_g\delta_m$, since lensing does not contain a bias term and both galaxy counts and lensing simply trace the matter distribution. However, in MG/DE theories the matter density appears coupled to the modified potential term, $\Sigma\delta_m$, which is thus clearly degenerate with the bias. This can inhibit the ability to constrain galaxy bias in large surveys, which can be a key source of errors. However, it is unlikely that viable galaxy bias models will have a profile $b_g(z) \sim \Sigma(z)$, so the detriment should be minimal for most models.

1.5 Galaxy Surveys

It is perhaps too obvious to be worth saying that, in order to determine cosmology from galaxy positions and shapes, we need a lot of them. Even theorists need to have a theoretically large number to subdue the noise terms. For decades we have

been observing increasingly large collections of galaxies, with greater precision, in order to supply the necessary quantities to deal with such finely graded statistics. The primary things which we need are positions on the sky (easy), redshifts (simple enough but time consuming), and ellipticities (both difficult and time consuming).

1.5.1 Redshift

The principle of redshift is well established in astrophysics, and is crucial to our understanding of the universe and its dynamics. By revealing the relative motion of an object and the expansion of space between us, we can understand how far away this object is in space and in time. The observation that, except for those closest to us (such as M31), all of the galaxies we observe are moving away from us led to the conclusion that the universe is expanding.

1.5.1.1 The cosmological redshift

The redshift is, from the theory side, easy to define. For a general metric $g_{\mu\nu}$, and for an emitter and receiver at fixed spatial coordinates (as is the case with 'fundamental observers' in cosmology i.e. objects which move with the Hubble flow $\dot{\chi} = 0$ to a good approximation), then the redshift is given by [26]

$$\frac{\nu_R}{\nu_E} = \frac{p_t(B)}{p_t(A)} \left(\frac{g_{tt}(A)}{g_{tt}(B)} \right)^{\frac{1}{2}}, \tag{1.189}$$

where $\nu_{R/E}$ are frequency of a photon at emission and reception, $p_t(A/B)$ are the time parts of four-momentum of the emitter (A) and receiver (B), and $g_{tt}(A/B)$ is the time-time component of the metric at the position of the emitter and the receiver. For cosmological redshift, we make use of the FRW metric

$$g_{\mu\nu} = \begin{pmatrix} c^2 & 0 & 0 & 0 \\ 0 & -a^2(t) & 0 & 0 \\ 0 & 0 & -a^2(t)F^2(\chi) & 0 \\ 0 & 0 & 0 & -a^2(t)F^2(\chi)sin^2(\theta) \end{pmatrix}, \tag{1.190}$$

where $F(\chi) \in \{\chi, \sin(\chi), \sinh(\chi)\}$. This metric is diagonal, so its inverse is also diagonal with components $g^{\mu\mu} = g_{\mu\mu}^{-1}$. Since photons follow null geodesics, where $\dot{\phi} = \dot{\theta} = 0$, our momentum equation reduces to

$$g^{\mu\nu} p_\mu p_\nu = \frac{1}{c^2} p_t^2 - \frac{1}{a^2(t)} p_\chi^2 = 0. \tag{1.191}$$

For null geodesics we also have $p_\chi = g_{\chi\chi} p^\chi = -a^2(t)\dot{\chi} = \text{const}$, so

$$p_t = a^{-1}(t) c p_\chi = a^{-1}(t) \times \text{const}, \tag{1.192}$$

and so

$$\frac{\nu_E}{\nu_R} = \frac{a(t)_R}{a(t)_E} = \frac{1}{a(t)_E}. \tag{1.193}$$

So if we detect light with a given wavelength, and we know the wavelength of the emission, we can calculate z and relate it to the scale factor, which is a function of time given by the dynamics of the cosmological model. (Note that if the scale factor $a(t)$ is not strictly increasing in t then there can be ambiguities. Luckily this is not the case in usual cosmologies up to the present day.)

1.5.1.2 High precision observation: spectroscopy

So given an observed galaxy, how do we determine its redshift? The most robust method is spectroscopy. By observing the spectra of galaxies, we can identify emission lines which have well known rest frame frequencies; it is then simply a matter of calculating the ratio of the emission line frequencies to their rest frame frequencies. These redshift estimates are known as *spectroscopic redshifts*, which is a combination of the cosmological redshift (the quantity in which we are primarily interested), peculiar velocity redshift, and redshift errors from the estimation itself. This process is our most reliable and sets our 'gold standard' for redshifts, with errors well below what is required to build reliable $n(z)$ estimates for cosmological inference. Unfortunately, this method is also time consuming, and collecting full spectra across a broad range of wavelengths for the vast numbers of galaxies required by cosmologists is simply infeasible at this time.

1.5.1.3 High volume observation: photometry

The other method available to us is photometry, for which the redshift estimates will be referred to as *photometric redshifts*. (For objects at redshift z, with spectroscopic redshift estimate z_s and photometric estimate z_p, we will generally have $z \neq z_s \neq z_p$, but $\langle |z_s - z| \rangle \ll \langle |z_p - z| \rangle$.) Photometric redshifts are calculated from much less information than their spectroscopic counterparts. Rather than using an entire spectrum to find emission lines, one uses only the intensity in five colour bands. This is an enormous reduction in information, and a variety of methods exist to try to overcome this barrier [63]. In general, empirical relations are found between the photometric parameters and the redshift by gathering photometric *and* spectroscopic information for a sufficiently large set of galaxies, and finding a fitting function between the observed photometry (input) and the spectroscopically calculated redshifts (output). A relationship from colours to redshift is found which attempts to minimise the redshift errors $\sum (z_p - z_s)^2$; nowadays this task is typically left to machine learning methods such as artificial neural networks [82][124]. Each of these photometric redshifts carries a significant intrinsic error, although it is possible to construct rough photometric redshift distributions.

1.5.1.4 Peculiar velocities

Our discussion of cosmological redshift is somewhat naïve. As we have mentioned before, it only applies to comoving emitters and receivers. For objects with significant peculiar velocities, the expression does not hold and the velocity of the emitter and/or receiver must be taken into account. The relative motion of the objects along the line of sight makes a contribution to the redshift, which (if ignored) distorts the redshift distribution of the galaxies. These distortions are known as 'redshift space distortions' (RSD). There are two principle effects with which we might be concerned:

- 'Finger of God' (FoG) effects: Galaxies within a cluster can have high velocities, which are randomly distributed in direction. This creates a distribution of redshift contributions stretching the cluster towards and away from us in redshift space (but not affecting their angular positions), creating the appearance

of an elongated filament pointing towards our position in space. [47]

- The Kaiser effect: Galaxies which are falling into an overdensity will have correlated motions towards the centre of infall. This means that the galaxies furthest from us (higher redshift) will have peculiar velocity towards us (a blueshifting distortion), and those closest (lower redshift) will have peculiar velocities away from us (a redshifting distortion). This results in a distribution which appears slightly compressed in redshift space. [48]

To linear order, redshift space distortions can be largely taken care of on a statistical level by modifications to the galaxy number count window functions $W_n(l,k)$ in the $C(l)$ formalism [49]. As we have discussed earlier, there is also redshift information in the LSS correlation functions; we shall see later how to exploit these to both obtain more precise redshift distributions and obtain less biased cosmological inferences.

1.5.2 Observing Ellipticity

Reliably observing galaxy ellipticity is an extraordinarily challenging task. It boils down to calculating the shape of a galaxy in an image, but this is not simple. Firstly, the images tend to be low resolution, maybe just a few pixels in each direction! This makes the shape recognition process very difficult, and with the volume of galaxies required it naturally must be automated. (As with photometric redshifts, this often means appealing to machine learning.) The issue is compounded by the fact that in weak gravitational lensing the shape changes to the galaxy are much smaller than the average intrinsic ellipticity. In order to get a grip on weak lensing, we needn't know a single source so well, but rather look for correlations among large sets of galaxies.

Given a sample of ellipticity measurements, one finds that if there are no preferred orientations of galaxies, then the mean of the observed ellipticity can be used as a measure of the reduced shear [31]:

$$g = \frac{\gamma}{1-\kappa} = \langle \varepsilon \rangle \tag{1.194}$$

The situation is more complicated when there are intrinsic alignments however, which is one of the most prominent problems facing weak lensing [32][33].

When analysing the shape of a galaxy image one must deconvolve the point spread function (PSF). The PSF is calibrated by looking at the shapes to foreground star images, whose properties are well understood and which are not lensed.

1.5.3 Relevant Surveys: the Dark Energy Survey and Euclid

The work in the first two chapters of this thesis (dealing with LSS) will be predominantly concerned with two major galaxy surveys: the Dark Energy Survey (DES), which is already underway and has released some early data; and the Euclid satellite which, hopefully, will soar into space around 2020 and not just blow up on the launch pad. These missions collect information, photometric and spectroscopic, on millions or, in the case of Euclid, billions of galaxies. This is supplemented in our theoretical work with CMB temperature information, which in practical terms would come from a survey such as Planck.

DES [139][140] is a ground based imaging survey, which will view galaxies over a patch of $5000\,\mathrm{deg}^2$ over 525 nights over a time scale of five years, and expecting to see a total of around 300 million galaxies in the wide-area survey. The camera itself, DECam, is a 519 megapixel optical CCD camera, which observes in four bands (which will provide the necessary information for photometry). DES has already begun releasing data, and the Year 1 cosmological results have been published [156][157]. Since DES is an imaging survey only, data will need to be supplemented from other sources if one wishes to include spectroscopy. BOSS (Baryon Oscillation Spectroscopic Survey) is a good candidate to combine with DES, and has provided a data set down to very low redshift. DES is expected to observe in the redshift range up to $z \sim 2$, so we will need more information in order to cover the redshift range with spectroscopy.

Euclid is an ESA satellite mission expected to launch in 2020, and is predicted to observe approximately 1.5 billion galaxies [142], and will cover an area of the sky at least $15000\,\mathrm{deg}^2$ up to a goal of $20{,}000\,\mathrm{deg}^2$. The total survey will take six years to complete, and thus will overlap with other next generation experiments

such as LSST. It will carry two detection instruments, VIS and NISP, which will carry out photometry in different bands. NISP will also measure spectroscopy when photometric filters are not applied, and thus provide a very significant spectroscopic sample of 50,000 galaxies. This sample will cover the range $0.7 \lesssim z \lesssim 2$, which is complementary to the BOSS selection which reaches $z \lesssim 1$. The fundamental physics goals of Euclid are very broad [143], and we shall be particularly interested in its power to constrain deviations from GR.

1.6 Computing and Cosmology

The calculation of cosmological quantities, from theory or inferred from data, is a computational discipline. Numerical methods are typically required to solve the majority of differential equations and calculate statistical estimators in all but trivial models. Cosmological inference, in particular, requires the calculation of quantities for many thousands of combinations of model parameters. In this thesis we shall work with two methods of inference: nested sampling, and machine learning. These fields are both vast in their own right, and we shall merely briefly review these methods as applied to this work.

1.6.1 Numerical Sampling

The most rudimentary way to find an estimate of a probability distribution is to sample it at regular intervals on a grid, and interpolate. Such naïve methods are only feasible with very few parameters, as the number of points which need to be calculated grows exponentially with the number of dimensions. When working with cosmology, where the number of parameters can easily be more than twenty, we require methods which obtain a reliable estimate of the probability distribution in a more efficient way. These are called sampling methods, with Markov Chain Monte Carlo (MCMC), Gibbs sampling, and nested sampling being popular choices. All seek to calculate the posterior distribution:

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}. \tag{1.195}$$

The quantity $P(D)$ – the probability of the data – is called the evidence (sometimes denoted Z). Nested sampling seeks to calculate the posterior distribution as well as an accurate estimate of the evidence. It is this quantity which is vital in comparing the viability of models with different parameterisations in a Bayesian way.

1.6.1.1 Nested Sampling

We shall use the ellipsoidal nested sampler PLINY throughout this thesis. Nested sampling is first described in a paper by Skilling [161], and perhaps the most widely used nested sampler in cosmology is MULTINEST [162]. (The successor to MULTINEST, POLYCHORD [163], is also a variant on nested sampling.)

The most important property of nested sampling, is that the evidence can be calculated as the sum of the posterior weights of the points in the chain:

$$Z = \sum_i w_i L_i, \tag{1.196}$$

where L_i is the likelihood and w_i the prior weight of each point. The points in the chain form a representative sample of the posterior distribution, and the posterior weight within a given volume can be approximated as the sum of the weights of the sampled points within that volume. The key to nested sampling is the way that the method navigates the prior volume, and estimates the prior mass of each sample which is accepted into the chain. The objective is to calculate a chain which is ordered in monotonically increasing likelihood and decreasing 'prior mass' X, where $X \in [0,1]$ is defined such that

$$X(L^*) = \int_{L(\theta)>L^*} \pi(\theta)\, d\theta, \tag{1.197}$$

where $\pi(\theta)$ is the prior, and θ stands for the parameters of the model i.e. X is the fraction of the prior occupied by points whose likelihood is greater than some threshold value L^*. As we increase this likelihood threshold, X decreases to 0 at $L^* = L_{\max}$. The evidence integral can be expressed as

$$Z = \int L\pi\, d\theta = \int L \frac{dX}{d\theta}\, d\theta = \int L\, dX, \tag{1.198}$$

which is easy to approximate numerically since L is monotonically decreasing with X by definition, and the integral is bounded, so this is generally well behaved and can be evaluated (including error estimates) using simple methods provided there is a reasonably fine sampling.

In nested sampling, one initially draws points uniformly from X, or (equivalently) from the parameter space θ according to the prior $\pi(\theta)$. The prior mass X_i of this point is estimated, and its likelihood recorded. The algorithm then proceeds by sampling uniformly $X_{i+1} \in [0, X_i]$, or equivalently from the parameter space according to the prior and provided $L_{i+1} > L_{\min}$ (i.e. inside the largest iso-likelihood contour). This may be achieved a number of different ways; the method of ellipsoidal sampling is briefly discussed in the next section.

Approximations for the prior mass X_i of each point are made statistically. In the simplest form, one can crudely estimate the factor t by which the prior mass is likely to shrink with each additional sample,

$$X_{i+1} = t_i X_i. \tag{1.199}$$

When N points are drawn, we must estimate the prior mass contained within the iso-likelihood contour defined by the lowest likelihood point. For a given shrinkage factor $t \in [0,1]$, the probability of containing $N-1$ other points inside is

$$P(N-1|t) = Nt^{N-1}, \tag{1.200}$$

since the minimum likelihood point can occur in any position in the list. Now we can say that

$$\langle t \rangle = \int tP(t|N-1)\,dt = \frac{N}{N+1}, \tag{1.201}$$

where we have used Bayes' theorem and the fact that $P(t) = P(N-1) = 1$, which can be seen from integrating equation 1.200 with respect to t, and from the fact that t is uniformly distributed within the unit interval.

1.6.1.2 Ellipsoidal Sampling

Ellipsoidal sampling [164] is an alternative to MCMC-style methods, which is employed in samplers such as PLINY and MULTINEST. In nested sampling, one typically draws N points from the available prior space. When we need to draw a new point, it must be done such that $L_{\text{new}} > L_{\text{min}}$. In order to do this efficiently, one method is to construct an ellipse which just encloses the current live points (the minimum spanning ellipse), enlarge it by some (usually pre-selected) factor f to account for the likelihood being not entirely ellipsoidal, and sample uniformly from inside this ellipse until one finds a likelihood which is sufficiently large. The success and efficiency of this algorithm rely to some extent on tuning the parameters N and f, but it is able to provide efficient and scalable sampling in many practical applications including cosmology.

1.6.2 Machine Learning with Artificial Neural Networks

Machine learning is a fast growing research area in both computer science and applied areas such as physics. We shall discuss briefly the field of machine learning and the operation of artificial neural networks (ANN), which will be used in Part II of this thesis.

1.6.2.1 Principles of Machine Learning

Machine learning is an approach to problem solving whereby algorithms attempt to match functions using statistical information and an iterative process in order to improve a model, rather than having an analytic model for each problem derived and hard coded. Goodness of fit is evaluated using a 'cost function' which penalises solutions which do not match the data well; the objective is thus to minimise the cost function. The algorithm must describe an efficient way to converge on an (almost) optimal solution. Approaches are generally non-deterministic, and often benefit from being run with multiple initialisations. Most approaches to machine learning are closely related to maximum likelihood methods, with implicit assumptions about the likelihood and priors being built into the cost function or structure and operation of the algorithm.

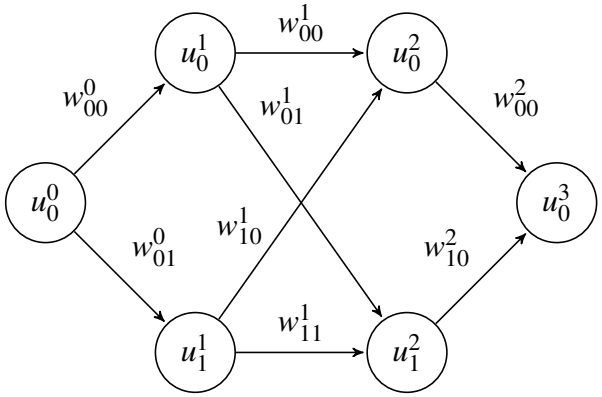
There are a wide variety of machine learning algorithms available which have been studied for decades; the choice of machine learning algorithm depends strongly on the necessities of the problem at hand. Signal analysis in image or sound processing may for example be well suited to algorithms which mimic functions using a linear combination of basis functions, with signals being built from sinusoids of different frequencies. Functions are then represented as a point in a vector space. In such methods, sparsity may be achieved by discarding irrelevant basis functions during the minimisation process, speeding up the algorithm. In other scenarios, selection of suitable basis functions may be difficult, and a more flexible representation of functions may be preferred. Artificial neural networks have excellent flexibility at the cost of obscurity, in the sense that the function parameters are not readily interpreted in an intuitive physical way.

1.6.2.2 A Simple Artificial Neural Network

A neural network is simply a mathematical function which is evaluated in stages and controlled by a set of parameters called 'weights'; it is usually represented as a graph. Each edge between nodes has a weight, and each node represents an 'activation function' which is included in a series of function compositions defined by the edges which connect them. It will be discussed further, including its application to physics, in chapter 4; here we shall provide a brief diagrammatic introduction to the construction of the function and the derivation of the weights.

The ANN graph may be structured in different ways; a feed-forward ANN is a directed acyclic graph, whereas other models such as Hopfield Networks are undirected graphs where all nodes are connected to all others (a clique) [165]. This thesis will make use of an implementation of a multi-layer perceptron (MLP), a kind of feed-forward network, called ANNZ [82]. Figure 1.2 represents a simple example of a such a graph structure. The ANN is composed of a series of layers of nodes. In the first layer we have the input(s), whilst the final layer is the output(s). In between, there are one or more 'hidden layers', which provide the complexity and flexibility of the function. Each of these layers may have different numbers of nodes. Each node is connected to all of the nodes in the following layer by an edge,

Figure 1.2: Diagram of a simple ANN with a single input and output, and two hidden layers with two nodes each. Note that the edges are directional, and represent the propagation of information through the network. There are four layers, indicated by superscripts: layer 0 is the input, layers 2 and 3 are the hidden layers, and layer 3 is the output. Node labels within each layer are denoted by subscripts, and thus each node is uniquely labeled u_i^k. We shall also use this to denote the value stored at this node. An edge weight from node u_i^k to u_j^{k+1} is labeled w_{ij}^k; edges may only connect nodes to nodes in the successive layer, not to nodes in the same of previous layers, and not skipping layers. The graph representation of the neural network is a diagrammatic representation of the function calculated by the ANN.



which carries a weight.

Given this structure, how is the output calculated from the input? Each node carries a value of its own, which is calculated from the nodes which precede it (with the exception of the input layer, whose values are simply the inputs themselves). For each node, the value there is

$$u_j^k = g\left(\sum_i w_{ij}^{k-1} u_i^{k-1}\right), \tag{1.202}$$

where the index k labels the layer of the network (with $k = 0$ the input), i and j are node labels within that layer, u_j^k is the value of node j in layer k, w_{ij}^k is the weight of the edge connecting node i in layer k to node j in layer $k+1$, and g is the 'activation function'. For example, in figure 1.2, the value u_0^1 (in the first node of the first hidden layer) is

$$u_0^1 = g\left(w_{00}^0 u_0^0\right), \tag{1.203}$$

and the value u_0^2 is

$$u_0^2 = g\left(w_{00}^1 u_0^1 + w_{10}^1 u_1^1\right), \tag{1.204}$$

$$= g\left(w_{00}^1 g(w_{00}^0 u_0^0) + w_{10}^1 g(w_{01}^0 u_0^0)\right). \tag{1.205}$$

The final output is the value u_0^3, which is determined likewise. One can see that as the network increases in size, these functions will become cumbersome to write explicitly and difficult to scrutinise analytically given a set of weights. In this small network we have already eight weight parameters which are free to vary; a practical network may have dozens of nodes and hundreds of weights.

1.6.2.3 Training and Convergence

Operation of the ANN requires three different sets of data: the training set, the validation set, and the testing set. (The testing set is not strictly necessary for training, but is used to evaluate the success of the regression.) Each set contains a list of inputs $\vec{x}$ paired with a list of target outputs $\vec{y}$ (i.e. the 'correct' outputs of the function). The objective of the ANN is to derive a function which, from the inputs, produces a value as close as possible to the target outputs. (This is known as 'supervised learning', because we provide the algorithms with examples of the output we would like it to produce.) These sets must contain entirely separate data with no points appearing in more than one set. The cost function used for the optimisation of the weights is calculated from the data in the training set, whereas the convergence criterion is based on the validation set; this is to avoid overfitting to the data in the training set. During training the cost function for the training set should continue to drop whereas the cost function for the validation set should plateau; the algorithm is considered converged when the change in fit to the validation set becomes sufficiently small or if the cost for the validation set starts to increase due to overfitting. The testing set can then be used to calculate statistics which evaluate the goodness of fit of the ANN function as an independent set not used in the training process at all.

1.6.2.4 Calculating the Weights

For a given network structure (graph), which fixes our functional form, we are left with the task of determining the optimal set of weights $\vec{w}$ given the data. For this we need two things: a definition of optimal, and an algorithm for getting there. Optimal is determined by a cost function, which is formed of two terms. For a set of inputs $\vec{x}$ and target outputs $\vec{y}$, and an output function F_{ANN} from the ANN,

$$C = \frac{1}{2}\sum_d |F_{\mathrm{ANN}}(x_d) - y_d|^2 + \beta \sum_{i,j,k} \left(w_{i,j}^k\right)^2, \tag{1.206}$$

where the index d runs over data points, and i, j, k are edge labels. Minimising the first term results in a simple least squares approach. The second term is called a regulariser, and its purpose is to avoid extreme weights. A solution is deemed optimal if it minimises this cost function. Although this is a standard choice (in ANN and also other methods), it is not the only one that one could make in principle, and the definition of the cost function will have profound effects on the behaviour and efficacy of the technique. A cost function such as this would be expected to perform well when results are analysed with, for example, the root-mean-square scatter as a metric (due to its similarity to a least squares cost function), but may not perform well under other statistical measures.

Given our definition of optimal, the weights now need to be varied so as to seek out the minimum of the cost function. In general, algorithms will converge to a local minimum (since to be sure of finding a global minimum one would have to explore the entire space, which is exactly what we must avoid). Weights are randomly initialised, and thus repeated calculations with different random seeds may yield different local minima. The minimisation may proceed by a number of methods; the ANN software used in this thesis, ANNZ, uses a quasi-Newtonian method. This method, described in [166] (which we follow here), expands the cost function around the current weight vector $\vec{w}$:

$$C(\vec{w}) = C(\vec{w}_0) + (\vec{w} - \vec{w}_0) \cdot \nabla C(\vec{w}_0) + \frac{1}{2}(\vec{w} - \vec{w})^T \mathbf{H} (\vec{w} - \vec{w}_0) + ..., \tag{1.207}$$

where the gradient is taken w.r.t. the weights, and **H** is the Hessian matrix, defined as

$$H_{ab} = \frac{\partial^2 C(\vec{w}_0)}{\partial w_a \partial w_b}. \tag{1.208}$$

Note that here the weights are labeled by a single index which simply indicates their position in the vector $\vec{w}$; the specific ordering of the weights does not matter. The minimum of the cost function is found by seeking the stationary point

$$\nabla C(\vec{w}^*) \approx \nabla C(\vec{w}_0) + \mathbf{H}(\vec{w}^* - \vec{w}_0) = 0, \tag{1.209}$$

where $\vec{w}^*$ indicates the weight at the (approximately) optimal point. By rearrangement

$$\vec{w}^* = \vec{w}_0 - \mathbf{H}^{-1}\nabla C(\vec{w}_0). \tag{1.210}$$

The gradient of the cost function may be calculated analytically; the inverse of the Hessian will likely have to be numerically approximated. This process may be applied iteratively, with each $\vec{w}^*$ becoming the $\vec{w}_0$ for the next iteration. Or, put more explicitly,

$$\vec{w}_{n+1} = \vec{w}_n - \mathbf{H}_n^{-1}\nabla C(\vec{w}_n). \tag{1.211}$$

The calculation of the gradient may be performed in the same was as described in back-propagation methods. Given the sigmoid activation function

$$g(x) = \frac{1}{1+e^{-x}}, \tag{1.212}$$

one finds that

$$g'(x) = g(x)\left[1 - g(x)\right]. \tag{1.213}$$

When the cost function is differentiated we have

$$\frac{\partial C}{\partial w} = \sum_d \frac{\partial F_{\mathrm{ANN}}(x_d)}{\partial w}(F_{\mathrm{ANN}}(x_d) - y_d) + 2\beta w, \tag{1.214}$$

where we have suppressed the indices on w to avoid clutter; w nevertheless rep-

resents a specific weight w_{ij}^k. For what follows we shall focus on a single datum, which might nevertheless be a vector if there is more than one output. The value of the function F_{ANN} is just the value of the output node(s) u_j^K, where K is the total number of layers. Differentiating by a particular weight w_{ab}^c and ignoring the regulariser (which is trivial), one finds that

$$\frac{\partial C}{\partial w_{ab}^c} = u_j^K(1-u_j^K)\,[V]\,(u_j^K - y_j), \tag{1.215}$$

where

$$V = \frac{\partial}{\partial w_{ab}^c}\sum_i w_{ij}^{K-1}u_i^{K-1} = \begin{cases} u_a^{k-1} & c = K-1, b = j \\ \sum_i w_{ij}^{K-1}\frac{\partial}{\partial w_{ab}^c}u_i^{K-1} & c \neq K-1. \end{cases} \tag{1.216}$$

In the latter case, the derivative of the node value u_i^{K-1} is calculated in the same way; in this fashion, we can calculate the derivative with respect to any weight by propagating back through the network via all paths which include the edge to which that weight pertains, using only the values of nodes and weights which have already been calculated anyway. This allows for efficient analytic evaluation of the derivative of the cost function with respect to the weights.

1.6.2.5 Connection with Probability Theory

As a final note on ANN methods, we shall look at the consequences of this choice of cost function in terms of a likelihood analysis, and thus connect machine learning to more traditional inference. Naturally from Bayes' theorem we have

$$P(\vec{w}|\vec{y}) \propto P(\vec{y}|\vec{w})P(\vec{w}), \tag{1.217}$$

where $\vec{y}$ are the targets for an input vector $\vec{x}$, and we are interested in the posterior probability of the weights since these fully determine the solution once a network structure has been selected. If we assume a Gaussian likelihood for the data with unit variance and mean $F_{ANN}(\vec{x})$, and a Gaussian prior on $\vec{w}$ centred on $\vec{0}$, then we

arrive at

$$P(\vec{w}|\vec{y}) \propto e^{-\sum(F_{\rm ANN}(x_d)-y_d)^2/2} e^{-\sum w_i^2/2\sigma^2}. \tag{1.218}$$

Taking logs we find that

$$\log\left[P(\vec{w}|\vec{y})\right] = -C + \text{const}, \tag{1.219}$$

where C is the cost function, and $\beta = \frac{1}{2\sigma^2}$. The variance of the likelihood can be altered by multiplying the first part of the cost function by a constant α. In the absence of regularisation, this has no impact on the maximum likelihood / minimum cost point. With a prior included, all that matters is the relative variance of the likelihood and the prior, so the same result will be achieved with unit variance on the likelihood and replacing $\beta \to \frac{\beta}{\alpha}$, and thus we need only control one parameter. Given this relationship, maximising the log-posterior (and therefore maximising the posterior) is equivalent to minimising the cost function as long as:

- the likelihood of the data given the model may be taken as Gaussian,
- the priors on the weights are Gaussian and centred on zero.

This connection with probability theory provides some motivation for this choice of cost function, but other cost functions may be derived from the use of other likelihoods and priors if desired.

Part I

The Universe on Large Scales

Chapter 2

Joint Analysis of Cosmological Parameters and Photometric Redshifts

Stars, in your multitudes,
Scarce to be counted,
Filling the darkness
With order and light

Les Misérables

2.1 Introduction

Galaxy surveys have become in recent years an important source of data for cosmology, particularly for late time effects such as dark energy. Calculations of predicted statistical properties for a given cosmological model require the redshift distribution of observed galaxies to be known accurately. Spectroscopy has long been used to calculate accurate redshifts for objects, but this is a time intensive process requiring both detailed observation across the object's spectrum and careful analysis. In order to collect data for the vast numbers of galaxies required, current and future surveys are necessarily dependent on photometric redshifts for the majority of objects.

When photometry is used, approximate redshifts are calculated from a small number of intensities measured in (typically around five) broad bands. Standard

methods of inferring redshifts from photometric data are to use machine learning methods, such as artificial neural networks, or to fit template functions (see e.g. [63] for a review). These require that we have large training sets of galaxies for which we have spectroscopic redshifts. Additionally, the spectroscopic set must be representative of the full photometric set (in terms of both redshift range *and* the nature of the objects contained within the sample) in order to reduce both the error and the bias in the derived relation ([124], [36]). Unfortunately, the spectroscopic sample is rarely as large as we would like, and is even less often fully representative of the redshift range we wish to look at. Spectroscopic samples tend to be dominated by bright objects which are easier to study; a lack of spectroscopic objects, particularly at the extremes of the redshift range, tends to lead to larger errors in the redshift distributions reconstructed from machine learning techniques. (Often the middle of the range is reconstructed comparatively well, while the outer regions suffer.) Other methods may be more successful at the lower or higher end of redshift, but few methods can be confidently used across the entire range ([36]). In addition to this, the errors associated with any such reconstructions are large, and those who have attempted to reconstruct redshifts from photometric data will be familiar with the significant scatter around the spectroscopic redshifts (e.g. [37]).

Due to the significant inherent uncertainties in such redshift estimates, for analysis objects may be grouped into bins of similar redshift. In order to achieve the precisions desired for current and future generations of cosmological experiments, we need to be able to determine the redshift distribution of each of these bins with greater accuracy than has been possible by simply using standard fitting to spectroscopic data. The impact of the redshift distributions on cosmology, and the importance of knowing them to a high degree of accuracy, has been the subject of a number of studies such as [45], [54]. These suggest that to achieve the desired precision and accuracy in upcoming experiments such as LSST, we require that the mean and width of redshift bins to be known to $O(10^{-3}(1+z))$. [54] proposes that this may be achieved by calibrating the the photometric redshifts using information from cross-correlations with spectroscopic data.

Since the proposal of these ideas, there have been some studies looking into the potential for using cross-correlations for estimating photometric redshift distributions (e.g. [53], [59], [60], [51], [52]), as well as potential problems such as contamination as in [38]. These tend to focus on recovering the redshift distribution from simulations by comparing the correlation between some photometric data set and a spectroscopic sample at known redshift, and assuming some fixed cosmology. In the case of a practical analysis however, we will not know the cosmological parameters (the determination of which is, after all, the objective of such calculations), and the calculation of theoretical correlation functions is cosmology dependent. It is well known that the cosmological parameters and redshift distribution are degenerate, and hence we cannot estimate how well the redshift distribution can be constrained without also varying the cosmology itself, as uncertainties from the cosmology may become a significant factor. This may be particularly important where the region of overlap between the photometric sample and the spectroscopic sample is relatively small. Hence, in order to avoid biases or overly optimistic estimates of our constraining power, we must determine the cosmology and the redshift distribution together, rather than treating them as independent problems. Previous work such as [54] and [52] also use estimators which may be prone to finding local maxima, and do not explore the space as fully as a nested sampling approach using a full likelihood.

A significant amount of attention in recent years has been placed on the power of cross-correlations as a statistical tool for cosmology (such as [57], [46]). Using cross-correlations to calculate both redshift distributions and cosmological parameters implies that we may include these effects into one framework with relative ease. We demonstrate such a technique for calibrating the photometric redshift distribution from an initial estimate using a joint likelihood analysis with cosmology using the angular power spectrum $C(l)$. We take the errors in photometric redshift modelling into account by allowing the mean and width of the photo-z bins to vary as free parameters, just as we do with cosmological parameters. The redshift binning is then marginalised over in order to obtain the probability contours for cosmolog-

ical parameters. This allows us to study the extent to which we can constrain the photometric redshift distributions and simultaneously explore the impact of this information on cosmological inferences, in a manner which automatically treats the errors in the distributions in a Bayesian way.

In this chapter we investigate the extent to which photometric redshift bins can be constrained by cross correlations, and the impact on cosmological parameter inference in the case of large scale structure. We present a simplified experiment where we vary the width and mean of gaussian redshift bins, although we explain how the framework may be applied to higher moments also. For computational simplicity, only three photometric and three spectroscopic bins will be used (although this can be extended to fuller surveys at the expense of computation time); this simple model should suffice to demonstrate the power of the technique, as well as the degeneracies between the parameterisations of the redshifts and the cosmological parameters. The impact on future optical surveys will be greater though, as the same technique can be used to constrain photometric samples in weak lensing analyses, which may be used in conjunction with galaxy number counts to infer cosmology.

2.2 The C(l) Calculation

The angular power spectrum is split up into correlations between different bins and cosmological probes; the full object we wish to look at is $C^{ij}_{\alpha\beta}(l)$ where i, j vary over labels of bins and α, β vary over cosmological probes such as galaxy number counts or shear measurements. (Indices may be suppressed when they are not relevant.) $C^{ij}_{\alpha\beta}(l)$ is symmetric under simultaneous exchange of i, j and α, β.

2.2.1 The C(l) formalism

Following the approach of [56] – and later [39], [62] – for a particular probe of our cosmology observed projected on the sky in the direction of a unit vector $\vec{n}$, $X(\vec{n}) = \bar{X} + \Delta X(\vec{n})$, we may decompose the variation in this parameter ΔX into

spherical harmonics as

$$\Delta X(\vec{n}) = \sum_{l>0} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\vec{n}). \tag{2.1}$$

We may calculate the coefficients a_{lm} by using the orthogonality of spherical harmonics ($\int Y_{lm} Y^*_{l'm'} d\Omega = \delta_{ll'}\delta_{mm'}$) :

$$a_{lm} = \int \Delta X(\vec{n}) Y^*_{lm}(\vec{n}) d\Omega. \tag{2.2}$$

The $C(l)$s are defined from these coefficients by the relation:

$$C(l) = \langle a_{lm} a^*_{lm} \rangle. \tag{2.3}$$

In our case we are interested in the galaxy distribution as a tracer of matter; this is calculated from the data by analysing number counts across the sky. It is important to note that this does not require knowledge of $n(z)$: we do not use redshift information in calculating the angular power spectrum from the data.

For the theoretical modelling however, we do require knowledge of $n(z)$, as we must calculate the full power spectrum $P(k,z)$ which is then projected onto the sky. This projection, as we shall later see, is strongly dependent on z. To calculate the $C(l)$s we use the following equation ([62]):

$$C^{ij}_{\alpha\beta}(l) = \frac{2}{\pi} \int W^i_\alpha(l,k) W^j_\beta(l,k) k^2 P(k) dk. \tag{2.4}$$

The redshift distributions $n^i(z)$ enter the $C(l)$s through the window functions.

2.2.2 Window functions

Window functions allow us to project the distribution of galaxies onto the sphere and decompose into spherical harmonics. Here we will discuss only the window function for galaxy clustering, since we have not used other probes in this particular work.

2.2.2.1 The galaxy clustering window function

[44] derive a calculation for galaxy clustering information of the following form,

$$W_{\rm g}^{i}(l,k) = \int b_g(k,z) n^i(z) j_l(k\chi) D(z)\, dz, \tag{2.5}$$

where $n^i(z)$ is the redshift distribution in bin i, $b_g(k,z)$ is the galaxy bias, $D(z)$ is the growth function, and $j_l(k\chi)$ is the order l spherical bessel function. Note that the comoving distance to an object is a function of redshift $\chi(z)$.

2.2.2.2 Including Redshift Space Distortions

Redshift Space Distortions (RSD) are alterations to the redshift of a galaxy due to its peculiar velocity. This leads to a distortion of the galaxy distribution if we attempt to reconstruct the three dimensional information, with galaxies with peculiar velocity toward us appearing closer (at lower redshift) and galaxies with peculiar velocity away from us along the line of sight appearing further (at higher redshift). Since these peculiar motions are due to interactions with local gravitational potentials they contain cosmological information. RSD on linear scales can be included by an additional term in the window function, following [46].

$$\begin{aligned} W^{i}_{\rm RSD}(l,k,z) = \frac{d\log D}{d\log a} \int n(\chi) D(\chi) [& \frac{2l^2+2l-1}{(2l+3)(2l-1)} j_l(k\chi) \\ & - \frac{l(l-1)}{(2l-1)(2l+1)} j_{l-2}(k\chi) - \frac{(l+1)(l+2)}{(2l+1)(2l+3)} j_{l+2}(k\chi)] d\chi, \end{aligned} \tag{2.6}$$

the complete window function to be used in our $C(l)$ calculation is then given by the sum of these two terms,

$$W^{i}_{\rm LSS} = W^{i}_{g}(l,k) + W^{i}_{\rm RSD}(l,k). \tag{2.7}$$

In this chapter we will not consider the effects of galaxy bias, although some papers have noted the potential importance of evolving galaxy bias in determining redshifts from correlation data ([59], [60]). In the absence of a compelling bias

model however, bias may be best handled as nuisance parameter (or parameters) which is also marginalised over. This is demonstrated in [43], but for simplicity we choose a constant bias $b_g = 1$.

We can now see how the redshift distribution enters into the $C(l)$ formalism. If we have an accurate redshift distribution, for instance from a spectroscopic survey, then this is all we need to begin calculating our theoretical correlations. Unfortunately, photometric estimates are far from perfect, and photometric redshift errors if ignored may produce unforeseen effects in our computed $C(l)$s. We will now seek to understand what some of these effects may be.

2.2.3 The significance of n(z)

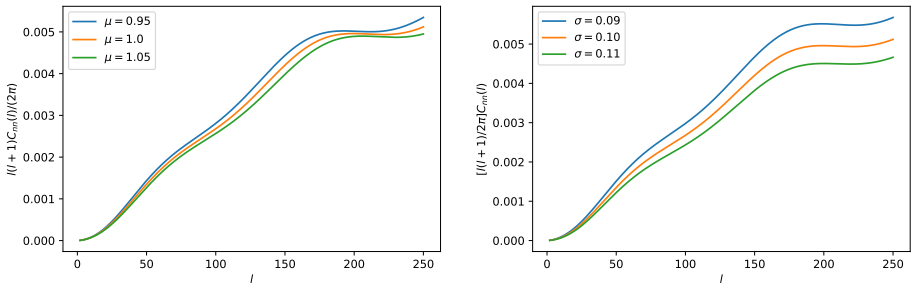
Intuitively one might expect that number counts on the same patch of sky will be highly correlated when close in redshift and less correlated when widely separated. If we have a spectroscopic sample and a photometric sample that overlap in redshift, then they will contain objects in the same larger clustering structures, which we will be able to see as boosts in their correlations. We do not expect to see clustering over very large distances, so we expect that samples widely separated in redshift will show very weak cross-correlations.

We can put this intuitive understanding on a more mathematical foundation. From the definition of the window functions and the $C(l)$ calculation, we can see how we expect $n(z)$ to affect our calculated $C^{ij}(l)$. If, for the sake of simplicity, we assume a k-independent bias $b_g(z)$, then the k dependence of $W(l,k)$ comes entirely from the spherical Bessel function $j_l(k\chi)$. The window functions oscillate as a function of k, made of contributions with different frequencies set by the spherical bessel functions in the integral. Hence the redshift range of the integral sets the range of frequencies present in the window function. If the distribution for a particular bin $n^i(z)$ is close to zero outside a particular range (for instance, if we model $n(z)$ as a top hat or Gaussian function) then the integral over z has a fairly small range which contributes significantly. If two bins $n^i(z)$ and $n^j(z)$ are separated in z by significantly more than their variance, then our two window functions $W^i(l,k)$ and $W^j(l,k)$ will have only very small contributions with the same frequency. The

product of two oscillating functions with different frequencies will tend to average to zero when integrated over, so we would expect the integral over these two window functions to be small. If, however, the redshift ranges overlap in regions of significant number density, then there will be significant contributions to both window functions with the same frequency and forms. These, when integrated over, will not average to zero and give a large contribution to $C^{ij}(l)$. Hence we expect the $C^{ij}(l)$s to be dependent on the amount of overlap between distribution functions in different bins, with significant overlaps in areas with high number density giving the strongest signals.

In addition to the overlap between bins, the spherical bessel function in equation 2.5 also tells us more about the redshift dependence of $C^{ij}(l)$: we expect stronger signals from distributions at lower redshift where the amplitude of $j_l(k\chi(z))$ is higher. Furthermore, there has been greater structure formation at low z, so $D(z)$ will be greater. So whilst the overlap between bins will determine the relative power in cross-correlations compared to auto-correlations, moving all the bins together up or down in redshift can shift the amplitudes of all the signals together. The redshift distribution is of course not the only thing which will affect our signal, and cosmological effects enter into our equations through the growth function ($D(z)$ in equation 2.5) and the power spectrum ($P(k)$ in equation 2.4). This is the source of a very important degeneracy between our redshift distributions and cosmological parameters, particularly those such as A_s or σ_8 which strongly control the amplitude of $P(k)$.

In order to fix the redshift distribution, we need bins which overlap our photometric redshifts but are strongly anchored so that any changes in photometric bins, even moving coherently, will be captured by the $C(l)$s. For this we require spectroscopic data, which is well known enough to have rigidly fixed $n(z)$, which overlaps our photometric bins. Note that our only criterion here is that our spectroscopic and photometric data overlap, and not - unlike with template and machine learning techniques - that the spectroscopic sample be an unbiased representation of the photometric sample.



(a) **(b)**

Figure 2.1: Left: Variation of an autocorrelation $C_{nn}(l)$ with the mean of the redshift bin μ (σ fixed at 0.1); Right: Variation of an autocorrelation $C_{nn}(l)$ with the width of the redshift bin σ (μ at 1.0). The power increases with decreasing μ since the there has been more time for growth to develop. Power decreases with increasing σ since samples spread over larger distances are less correlated than samples contained in a thin slice.

Fig 2.1 shows explicitly the effect that the properties of the redshift distribution have on the $C(l)$ function that we retrieve. These changes to the $C(l)$ can be degenerate with certain cosmological parameters, and thus any unknown or unaccounted for errors in redshift distribution can translate into biased and inaccurate cosmological inferences, as the cosmological parameters are forced away from their most likely values to compensate for the effects of errant redshifts.

2.3 Modelling the Redshift Distributions

2.3.1 Photometric redshifts

Due to the uncertainty in photometric redshifts, we cannot obtain an accurate redshift for each object that we have in our sample. We instead model a bin as a broader function which captures the distribution of redshifts which would be binned together.

In this work we will model photometric redshift distributions within a bin as a Gaussian distribution defined by their mean and variance,

$$n(z,\mu,\sigma) = G(z,\mu,\sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]. \tag{2.8}$$

In order to model the uncertainties in $n(z)$, we need to be able to control the

shape of the function in a quantitive way, ideally with as few parameters as possible. Each time we add a parameter, we are adding N_{bins} new dimensions to our parameter space to be explored by the sampler and hence our computation becomes exponentially more expensive. The most important parameters are the mean and width of the distribution; other adjustments to the shape can be abandoned without too much impact but nevertheless the method is general and in wider, strongly non-Gaussian redshift bins higher moments may be taken into account if necessary. When using the mean and variance of a gaussian distribution we may adjust μ and σ directly using the analytic formula for a Gaussian. To vary a general distribution, or to change the shape in other ways, you may refer to section 2.7.1.

Although this is the template used for all the bins in this study, we may also apply non-Gaussian distortions to these distributions to model more complex effects. It is also important to note that this method is by no means limited to Gaussian functions, and these may be easily replaced by an arbitrary function (with some parametrisation) $F(z, \mathbf{p})$, where $\mathbf{p}$ is the parameter vector to be marginalised over. In the simple example above $\mathbf{p} = (\mu, \sigma)$, although we may extend this to include skew and kurtosis, and have $\mathbf{p} = (\mu, \sigma, s, k)$ or some other vector of parameters. In a typical survey such as DES, photometric redshift bins have a standard deviation of approximately 0.1 to 0.2 ([50]); for our purposes we will take $\sigma = 0.1$.

2.3.2 Spectroscopic redshifts

In order to include spectroscopic redshifts into the same formalism, we model spectroscopic redshifts in bins with much narrower distributions. This will take the form of a much narrower Gaussian. (A narrow top-hat function may also be used, but smooth continuous functions are often computationally more stable.) We assume that spectroscopic information is known well enough that we do not vary these bins in the same way as the photometry, and so there is no parameter vector $\mathbf{p}$ to marginalise over. Spectroscopic bins will be modelled with a width of $\sigma = 0.025$, which requires spectroscopic redshifts to be estimated to within a few percent. Their thickness may be determined by the nature and quality of the spectroscopic sample, or as a compromise with computational efficiency. A small number

of wider bins is less computationally expensive than many narrow bins; the width of spectroscopic bins makes little to no impact on the length of computation (integrations are performed between fixed redshifts), however each additional bin adds two new parameters, which means that we have more integrations to perform (scaling as N^2_{bins}), larger covariance matrices, and a much larger parameter space which scales exponentially in volume with the number of parameters. Narrow bins allow us to look at very localised correlations at the cost of this additional computation.

2.4 The Likelihood Function and Sampling Methods

2.4.1 The likelihood function for C(l)s

The likelihood is calculated from $C(l)$ for each model (i.e. each parameter set) compared to the $C(l)$ calculated from the fiducial model. Let the fiducial model be known as model A, and the model we wish to investigate model B; we calculate a log-likelihood of seeing some fluctuations a_{lm} in the model B compared to the model A, and take then take an expectation value assuming the fiducial model A in the absence of any data, as described in [42],[35]. Since a_{lm} are stochastically generated, any given cosmology may generate a wide variety of a_{lm}, and each set of a_{lm} can give therefore give a different likelihood when compared against a model. Hence, with no reason to generate one particular set over another, one calculates the expectation value of these possible likelihoods, on the assumption that our a_{lm} were generated by the cosmology represented by A. This quantity is dependent only on the $C(l)$s calculated in each model, and the properties of the survey such as sky coverage and noise which remain constant throughout.

For measured a_{lm}, using the fact that the expectation value is zero, we have for a given cosmology X the relation

$$\text{Var}(a_{lm}) = \langle |a_{lm}|^2 \rangle_X = C_X(l) + N(l), \tag{2.9}$$

where noise is assumed to be isotropic and uncorrelated (shot noise) and taken into

account by the noise function $N(l)$. Assuming Gaussian distributions, we then have

$$P(a_{lm}|X) = \left(\frac{1}{2\pi(C_X(l)+N(l))}\right)^{\frac{1}{2}} \exp\left[-\frac{|a_{lm}|^2}{2(C_X(l)+N(l))}\right]. \tag{2.10}$$

We wish to calculate the (expected) likelihood function

$$\langle L\rangle = \left\langle \log\left[\frac{P(a_{lm}|B)}{P(a_{lm}|A)}\right]\right\rangle_A. \tag{2.11}$$

Given equation 2.10 we can write

$$\frac{P(a_{lm}|B)}{P(a_{lm}|A)} = \left[\frac{C_A(l)+N(l)}{C_B(l)+N(l)}\right]^{\frac{1}{2}} \exp\left[\frac{|a_{lm}|^2}{2(C_A(l)+N(l))} - \frac{|a_{lm}|^2}{2(C_B(l)+N(l))}\right], \tag{2.12}$$

and taking logs we obtain

$$\log\left(\frac{P_A}{P_B}\right) = \frac{1}{2}\log\left(\frac{C_A(l)+N(l)}{C_B(l)+N(l)}\right) + \frac{|a_{lm}|^2}{2(C_A(l)+N(l))} - \frac{|a_{lm}|^2}{2(C_B(l)+N(l))}. \tag{2.13}$$

We then take the expectation value assuming A using equation 2.9

$$L = \frac{1}{2}\left[1 - \frac{C_A(l)+N(l)}{C_B(l)+N(l)} + \log\left(\frac{C_A(l)+N(l)}{C_B(l)+N(l)}\right)\right]. \tag{2.14}$$

We must then take into account all of the a_{lm}, bearing in mind this expression is not dependent on m, so we have $2l+1$ identical terms for each l, and taking into account the fraction of the sky f_s observed.

$$L = \frac{f_s}{2}\sum_{l=2}^{l_{max}}(2l+1)\left[1 - \frac{C_A(l)+N(l)}{C_B(l)+N(l)} + \log\left(\frac{C_A(l)+N(l)}{C_B(l)+N(l)}\right)\right]. \tag{2.15}$$

This is the likelihood we will use for CMB temperature information ($C^{TT}(l)$) or when we calculate the autocorrelation of a bin. When we have multiple bins or cos-

mological probes where cross correlations must be taken into account, then we will have more than one $C(l)$ function for each cosmology. We then use a multivariate Gaussian distribution instead of simply the product of independent Gaussians. The covariance matrices are

$$\left[M_{X,l}\right]_{i,j} = C^{ij}_{(X)}(l) + \delta^{ij}N^{i}(l), \tag{2.16}$$

where X may be either A or B (with the relevant $C(l)$s calculated in the left hand side) and where $N(l)$ is the noise associated with the experiment. Noise is only added on the diagonal as shot noise between bins should not be correlated and hence not contribute to the covariance. This gives a probability distribution

$$P(\mathbf{a}_{lm}|X) = \frac{1}{\left((2\pi)^k|\mathbf{M}_{X,l}|\right)^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\mathbf{a}^T_{lm}\mathbf{M}^{-1}_{X,l}\mathbf{a}_{lm}\right], \tag{2.17}$$

for a $k\times k$ covariance matrix (i.e. cross correlating k bins). Repeating the above analysis, and using the following relations (where l,m subscripts have been suppressed for clarity, and we use summation convention over i,j)

$$\begin{aligned}\left\langle \mathbf{a}^T\mathbf{M}_X^{-1}\mathbf{a}\right\rangle_A = \left\langle a_i M^{-1}_{X,ij}a_j\right\rangle_A &= M_{A,ij}M^{-1}_{X,ij} \\ &= M_{A,ji}M^{-1}_{X,ij} = \left[\mathbf{M}_A\mathbf{M}_X^{-1}\right]_{jj} = \mathrm{Tr}\left[\mathbf{M}_A\mathbf{M}_X^{-1}\right],\end{aligned} \tag{2.18}$$

we arrive at the analagous log-likelihood to equation 2.15 for multiple $C(l)$s

$$L = \frac{f_s}{2}\sum_l (2l+1)\left[\mathrm{Tr}\left(I - M_{A,l}M^{-1}_{B,l}\right) + \ln\left(\det\left(M_{A,l}M^{-1}_{B,l}\right)\right)\right]. \tag{2.19}$$

It is easy to see that this is zero for A=B.

Table 2.1: Parameters for a Planck like CMB survey.

Band Frequency	70	100	143	217
Beam Width θ_b / arcsec	14.0	9.5	7.1	5.0
Noise Effective Temperature / $\mu K\sqrt{s}$	212	56	56	84
Detector Number $n_{\rm det}$	12	8	12	12

2.4.2 Noise parameters and survey assumptions

2.4.2.1 Galaxy number counts

We limit our model to shot noise, which is described by the noise function

$$N^i(l) = (\sigma^i(l))^2 = \frac{1}{\bar{n}^i} = \frac{4\pi f_s}{f_g^i N_g}, \tag{2.20}$$

where f_s is the fraction of the sky observed by the survey, f_g^i is the fraction of the total number of galaxies observed which lie within that redshift bin $n^i(z)$, and N_g is the total number of galaxies observed over the entire survey. For a DES like survery we assume that $N_g = 3\times 10^8$, $f_s = 0.12$ (from $A = 5000\text{deg}^2$), and a redshift range $0 < z \leq 2$ ([50]). Galaxy count $C(l)$s are calculated in the range $2 \leq l \leq 323$.

2.4.2.2 CMB TT information

Here we have a slightly more complex function which must take into account more survey information. We base our parameters on a Planck-like survey, based on the parameters described in [35]. Our noise function is

$$N_l^2 = \sum_{chan} \frac{1}{(\sigma_c\theta_b)^2} \exp\left(-\frac{l(l+1)\theta_b}{8\ln 2}\right), \tag{2.21}$$

$$\sigma_c = \frac{T_{\rm NE}\theta_{\rm sky}}{\sqrt{n_{\rm det}t}\theta_b}, \tag{2.22}$$

where θ_b is the beam width, $T_{\rm NE}$ is the noise effective temperature, $n_{\rm det}$ is the number of detectors, and t is the integration time assumed to be one year. We assume information is collected in four bands with parameters detailed in Table 2.1. For CMB information we assume $f_{\rm sky} = 0.65$; CMB-TT $C(l)$s are calculated in the range $2 \leq l \leq 3000$.

2.4.3 Computational details: UCLCl and PLINY codes

$C(l)$ calculations are performed using the UCLCL code developed at UCL by Filipe Abdalla, Sreekumar Balan, and myself, and the CLASS Boltzmann code ([77]) for the generation of the primordial power spectrum and transfer function. Within UCLCL most functions, including $n(z)$, are represented using splines. The spline representation is advantageous for this work because it allows us to easily manipulate and deform $n(z)$ in non-linear ways without having to define an analytic function with some parametrisation. (This means we could take an arbitrary form from, for instance, data and still manipulate it in the way described in this chapter.) We can vary the mean and variance for an arbitary distribution in a precise way. For the higher moments such as skew and kurtosis, we must vary these more heuristically for a general distribution, and these transformations may affect other moments. These may all be varied by applying transformations to the z variable of the $n(z)$ spline, as discussed in the appendix.

The nested sampling analysis is performed using PLINY (developed by Richard Rollins, Filipe Abdalla, and Sreekumar Balan), a nested sampler designed for parallel computation. It calculates a chain of points in the parameter space, with likelihoods and prior weights, and also outputs an evidence calculation. In order to calculate the posterior weight for each point in the chain we need to use Bayes' Theorem:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} \tag{2.23}$$

For all parameters in this analysis we assume flat priors with hard edges well away from the peak of the distribution. The evidence is not strictly necessary in this analysis as it is just a constant factor. The evidence is only required if we wish to perform a model comparison for models with different parameterisations.

2.4.4 The fiducial model

In this work we use a fiducial ΛCDM model. For the sake of computational efficiency, we take work only with flat cosmologies ($\Omega_k = 0$). We also restrict ourselves to varying seven cosmological parameters - $\{A_s, \Omega_\Lambda, \Omega_b, h, n_s, \tau_r, w_0\}$. Our fiducial

cosmology will be

$$\begin{aligned} A_s &= 2.5 \times 10^{-9}, \\ \Omega_\Lambda &= 0.7, \\ \Omega_b &= 0.06, \\ h &= 0.7, \\ n_s &= 0.95, \\ \tau_r &= 0.09, \\ w_0 &= -0.9 \end{aligned} \tag{2.24}$$

To speed computation we limit ourselves to flat cosmologies, and hence we will use $\Omega_{\rm cdm} = 1 - \Omega_\Lambda - \Omega_b$, which gives a fiducial $\Omega_{\rm cdm} = 0.24$. We use three photometric bins with mean

$$(\mu_1, \mu_2, \mu_3) = (0.8, 1.0, 1.2),$$

and standard deviation

$$\sigma_1 = \sigma_2 = \sigma_3 = 0.1.$$

For analyses with spectroscopy, we use three spectroscopic bins with mean

$$(\mu_1, \mu_2, \mu_3) = (0.7, 1.0, 1.3),$$

and standard deviation

$$\sigma_1 = \sigma_2 = \sigma_3 = 0.025.$$

Spectroscopic bins are assumed to be well known enough not to need variation in the nested sampling analysis, so this leaves us with an 13-dimensional parameter space (7 cosmological and 6 photometric binning parameters).

For our noise function, in this particular analysis, we have chosen DES like

parameters (described in section 2.4.2) with $f_s = 0.12$, $f_g^i = 0.2$ for all photometric bins, and $N_g = 3 \times 10^8$ (giving 6×10^7 galaxies in each photometric bin). For spectroscopic bins, we assume that we have 5×10^4 galaxies in each bin, roughly in keeping with the density in surveys such as BOSS or eBOSS ([61], www.sdss.org).

For the sake of reasonably rapid calculations we use a limited number of redshift bins in this demonstration, although most full surveys will use 5-10 photometric redshift bins. This will diminish our power to constrain the cosmological parameters somewhat due to a lack of information and coverage over much of the redshift range, but will be enough to demonstrate the power of the technique applied to the calibration of photometric redshifts.

2.5 Results

In this section we will present the results of our nested sampling analysis. We will first demonstrate the bias in cosmological parameters that is caused by having poorly estimated photo-z bins. We will then show what can be achieved using autocorrelations of photometric bins, where bins are allowed to vary freely; this will demonstrate where the degeneracies between photometric redshifts and cosmological parameters lie. Finally we will show results using cross correlations between both photometric and spectroscopic bins, which gives dramatically improved precision on the photometric bins, and we demonstrate the effect of this on the marginalised distributions for the cosmological parameters.

2.5.1 Cosmological parameter bias from n(z)

From equations 2.4 & 2.5, our theoretical prediction of $C^{ij}(l)$ is dependent on the redshift distributions $n^i(z)$ and $n^j(z)$. If we estimate properties of our redshift bins (in this case μ, σ) by fitting objects with known spectroscopic redshifts, then we will derive redshifts with some scatter around their 'true' value. These redshift errors have a knock-on effect on our inference of cosmological parameters. For example, if our estimated redshifts are too low, then theoretical power that we calculate will be too high; in order to match the observations, A_s may be lowered to match the power, and other parameters adjusted to get the best fit to shape. In this

section we will demonstrate such biases, and later we shall see how marginalising over redshift distributions can avoid them. Methods in estimating photometric redshifts often have an error in z of $O(0.1)$, which is large compared to what we would require to obtain precise results from a photometric survey. If we take the parametrisation obtained from this fitting on face value then we will reconstruct a slightly distorted $n(z)$. This means that when we fit our cosmological model, our cosmological parameters will inevitably be changed in order to counter the effect of the distortions in $n(z)$. We may analyse this case in our simple model by using the fiducial $n(z)$ for the "observed" $C(l)$s as described in section 2.4.4, but calculating our model $C(l)$s using bins fixed to have different parameters.

The cosmology used is the same as stated in section 2.4.4 but we shall only use two redshift bins at $\mu_1 = 0.8$ and $\mu_2 = 0.9$ to generate the fiducial $C(l)$s. When we attempt to recover the cosmological parameters with a sampling analysis, we use a fixed redshift distribution, biased with $\mu_1 = 0.75$ and $\mu_2 = 0.85$. The results are shown in Figure 2.2.

The bias is strongest in cases such as this where there is a systematic error causing the mean or standard deviation of bins to be consistently over or under estimated. In order to avoid this, we must reduce our reliance on fixed redshift distributions with large errors. In lieu of a method for sufficiently accurate redshifts from photometry, we must rely on marginalising in a bayesian framework, the results of which are described in the following sections.

2.5.2 Autocorrelations with photometric redshift bins

The simplest analysis that we can do is to use only our photometric redshift bins, and to only take into account autocorrelations. We will see that this means ignoring a great deal of information, and our bounds on cosmological and binning parameters are wide. Although in this case we are not taking into account the full information available to us, it is worth looking into since it is much less computationally expensive, and previous studies have been focussed on autocorrelations. We promote the mean and standard deviation of our redshift bins to fully independent parameters for our nested sampling analysis, allowing them to vary freely so that they can

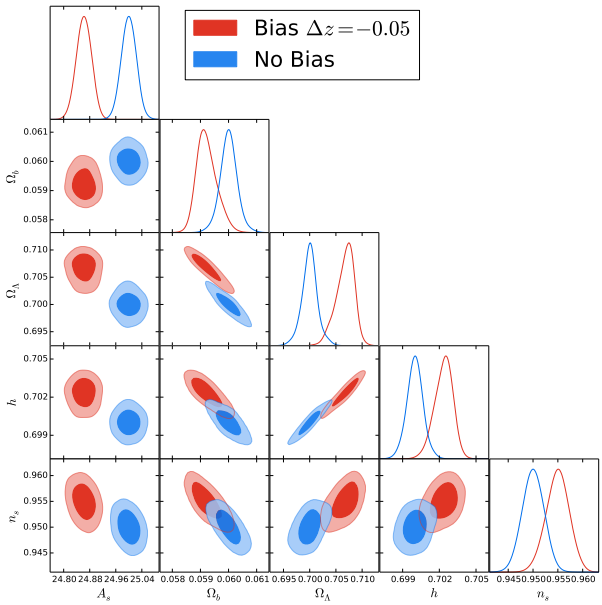


Figure 2.2: Probability contours obtained from the true redshift distribution (shown in blue) and from a biased redshift distribution (shown in red). A five parameter cosmology is derived from two photometric bins, where the fiducial cosmology and the blue contours use $\mu_1 = 0.8$ and $\mu_2 = 0.9$, whereas the red contours are derived on the incorrect assumption that $\mu_1 = 0.75$ and $\mu_2 = 0.85$ i.e. photometric redshifts are systematically underestimated. The blue contours are, by construction, centred on the fiducial parameters, whereas the red contours end up far from the fiducial parameters in order to compensate for effects in the $C(l)$ signal introduced by photometric systematics.

Table 2.2: 68% confidence ranges for inferred cosmological parameters using only auto-correlations, and using cross correlations with spectroscopy.

	$A_s \times 10^{10}$	Ω_b	Ω_w	w_0	h	τ_r	n_s
Fiducial	25	0.06	0.7	-0.9	0.7	0.09	0.95
Photometric	25.1 ± 1.5	0.060 ± 0.007	0.693 ± 0.033	-0.89 ± 0.05	0.697 ± 0.019	0.091 ± 0.029	0.950 ± 0.004
Photo $\times$ Spec	25.0 ± 0.4	0.060 ± 0.002	0.700 ± 0.008	-0.90 ± 0.02	0.700 ± 0.006	0.091 ± 0.009	0.950 ± 0.004
Fixed Redshift	25.0 ± 0.3	0.060 ± 0.002	0.700 ± 0.008	-0.90 ± 0.02	0.700 ± 0.005	0.090 ± 0.008	0.950 ± 0.004

be marginalised over. Here we use the three photometric bins described in section 2.4.4. This information is combined with CMB TT information in order to help constrain A_s, which is a problematic parameter in this analysis since it is extremely strongly degenerate with the standard deviation of a bin (see Figure 2.3). (It is also degenerate with the bias of the galaxy sample in a bin, which we have not investigated here; we discuss this to some extent in section 2.6.1.)

Despite the fact that we are not using any cross-correlations between bins, we can see the parameters for different bins are degenerate. This is because of the effect

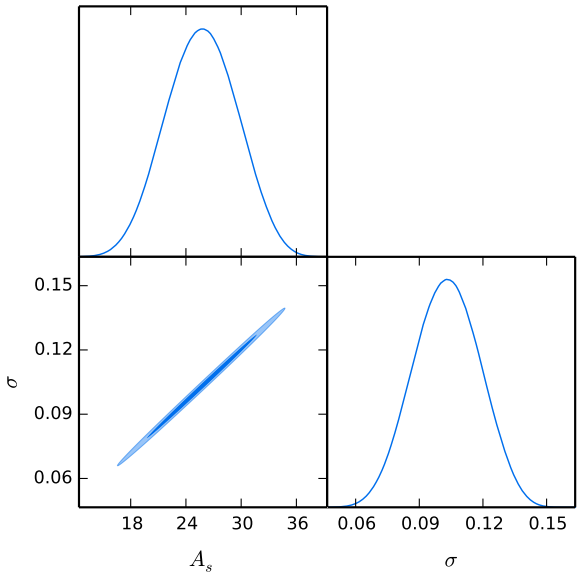


Figure 2.3: Probability contour obtained when varying only A_s and σ (the width of one photometric redshift bin), with $\mu = 1.0$ and all other parameters at their fiducial values, demonstrating the high degeneracy between photometric bin width and the A_s parameter.

Table 2.3: 68% confidence ranges for inferred photometric redshift bin parameters, using only autocorrelations and using cross correlations with spectroscopy.

	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3
Fiducial	0.8	1.0	1.2	0.1	0.1	0.1
Photo+Auto	0.794 ± 0.036	0.992 ± 0.045	1.191 ± 0.054	0.101 ± 0.005	0.101 ± 0.005	0.101 ± 0.005
Spec+Cross	0.800 ± 0.003	1.000 ± 0.003	1.200 ± 0.003	0.100 ± 0.001	0.100 ± 0.001	0.100 ± 0.001

that μ and σ have on the $C(l)$. For example, when all the bins are moved in the same direction, the effect is largely to raise or lower the power in each autocorrelation; this can be compensated for by adjusting A_s and other cosmological parameters appropriately. If however some bins are moved up in redshift, and some down, then the cosmological parameters struggle to compensate for the competing effects. The same is true for σ. This means that these parameters are constrained to move together to some extent.

We note that the degeneracy between σ and A_s is being prevented from exercising its full effect because A_s has been constrained significantly by the CMB information. Nevertheless A_s is strongly degenerate with τ_r (the optical depth at reionisation) and we note that τ_r is not well constrained in this instance. Although its effects have been mitigated by its constraint, it is still clear that there is a de-

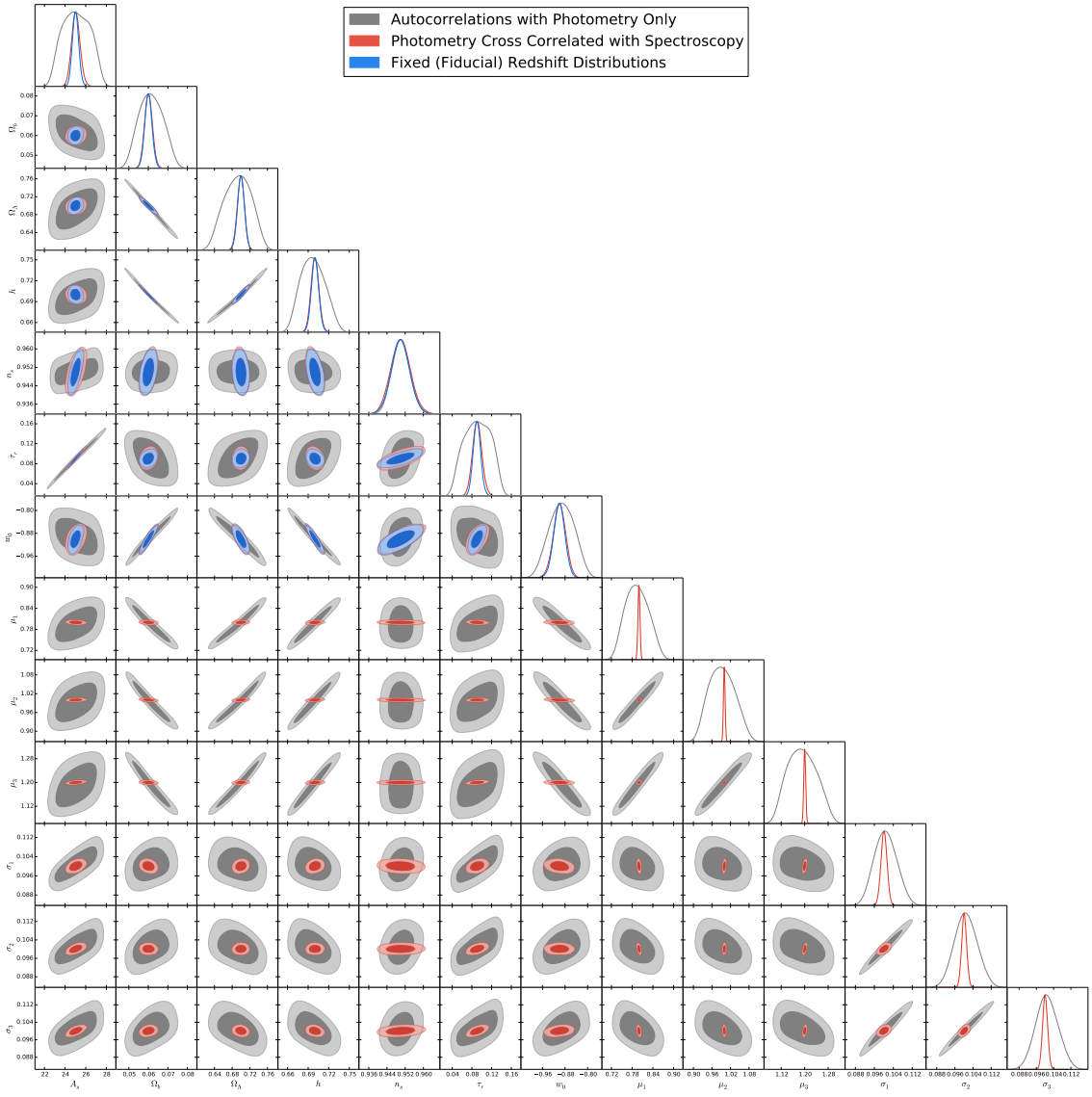


Figure 2.4: Probability contours obtained for cosmological and photometric binning parameters. Results from using autocorrelations of photometric bins only are shown in grey, and results from using photometric bins cross correlated with each other and spectroscopic bins are shown in red. For comparison, results using redshift bins fixed at the fiducial values are shown in blue. Results show clear improvements on all the binning parameters, as well as most cosmological parameters (with the exception of n_s), and contours using cross correlations between photometric and spectroscopic samples yield results very close to those with no redshift error.

generacy between σ of each bin and all of the cosmological parameters except n_s. Likewise μ is strongly degenerate with Ω_b, Ω_Λ, h, and w_0, and is not constrained up to the hard limits of the prior. This means that errors in estimates for binning parameters can propagate into cosmological parameters in a significant way.

We can understand the degeneracy between the cosmological parameters by considering their effects upon the $C(l)$s. The effect of σ is primarily to change the height of the $C(l)$s, which creates its degeneracy with A_s. Likewise, we know that moving μ to low redshift boosts power; since Ω_Λ and h suppress structure formation, these need to be lowered and Ω_b raised to get the C(l)s to match the fiducial model.

2.5.3 Cross correlating with spectroscopic redshifts

In this section we demonstrate the improvement attainable by cross-correlating with spectroscopic redshift data. Because of the overlap with spectroscopic data, we can show that the properties of the photometric redshift bins are now tightly constrained, and the degeneracies between bins are less pronounced. In most cases, the binning parameters cannot vary widely enough to have a noticeable impact on the cosmological parameters compared to the uncertainty already present.

We can see that there is increased precision in the cosmological parameters (except for n_s, which is almost entirely determined by CMB information here), with most bounds improving by a factor of two or more (Table 2.2). The slight widening in the posterior distribution for n_s which can be seen in Figure 2.4 is most likely due to the additional noise introduced to the galaxy clustering likelihood by looking at larger numbers of bins. Since galaxy number counts do little to constrain n_s at this level, and bins which are widely separated in redshift may produce correlation functions that are largely noise dominated (since they should be close to zero), this small addition of noise to the likelihood causes some spreading of this parameter. This could be tackled by ignoring widely separated bins if necessary (this would also speed up the likelihood calculation by reducing the number of integrations). Binning parameters (μ_i, σ_i) have been particularly tightly constrained, allowing us to know their values to percent level or better (Table 2.3). Further constraint can be

imposed upon them by having more spectroscopic bins to cover a greater fraction of the photometric redshift range, at the cost of computation time.

2.6 Discussion

From the results presented in section 2.5, we can see that different aspects of the analysis provide distinct benefits. The variation of $n(z)$, and its subsequent marginalisation, is essential for the removal of the bias from cosmological inferences. In order to utilise the full power of the $C(l)$ formalism, we must include cross-correlations as well as the well studied auto-correlations; these not only provide us with with much more information (improving our constraining power), but also help us to pin down the relationships between different photometric redshift bins more accurately. Since the photometric redshift parameters display a degeneracy with almost all of the cosmological parameters, it is crucial to have these distributions known as well as possible.

By using a theoretical likelihood and nested sampling approach this study is less idealised than most previous works such as Fisher matrix analyses; nevertheless the assumption of gaussian bins, and the small number of redshift bins, are simplifications that should be addressed in future work. As the analysis is extended with more bins, the computational complexity will increase significantly; the number of cross-correlations to calculate will increase as $O(n_{bins}^2)$ and the dimensionality of the parameter space as $O(n_{bins})$. It is possible to simplify in such cases by only considering cross-correlations between bins which are sufficiently close together (leading to a band diagonal $C^{ij}(l)$ matrix), since cross-correlations between widely separated bins will contain comparatively little information. Despite the computations intensity, as long as numerical errors remain tamed, we expect that increasing the number of bins and the density of the spectroscopic sample will improve the results. A survey such as Euclid should have a great deal of power to jointly constrain the redshift distribution and the cosmology with minimal disturbance to the confidence intervals for cosmological parameters.

Errors on redshift binning parameters are now $O(10^{-3})$, even with such sparse

spectroscopic data as we have simulated. Errors on the means of photometric bins are at ± 0.003 and errors on the width of bins is at ± 0.001 (see Table 2.3). This is extremely promising for future experiments, providing the possibility to extract reliable and precise cosmological parameters. As we look towards future experiments such as Euclid and LSST, and even with data currently being released from DES, a major focus in cosmology will be the nature of dark energy. The ability to distinguish between a cosmological constant, scalar field theory, modified gravity, or more exotic forms still, will be dependent on having well known redshift distributions, as can be seen by the strong degeneracy between the mean of redshift bins and the parameters Ω_Λ and w_0. A bias from improperly calibrated photometric data could easily generate a spurious result. Using this method will help to ensure robust analyses for current and future experiments.

This method can be applied to any $C(l)$ signal using photometric redshift bins (such as weak lensing or galaxy clusters) to calibrate their photometric redshift distributions. This means that when applied to future optical surveys, it will be able to benefit much more powerful analyses than the one outlined in this chapter, including a larger number of redshift bins, and a combination of signals from different cosmological probes. If the same photometry is used for both number counts and lensing, then both of these cross correlations will contribute to constraining the photometric parameters, as well as constraining the cosmology itself. It will be necessary for future observational work to extend this to non-gaussian distributions, including higher order moments or more generic spline models of $n(z)$, in order to model our observed photometric redshifts as best we can. When combined with lensing information, this technique can be applied to achieve improved results in Modified Gravity or Dark Energy studies, where biases can lead to spurious detections and high precision is needed.

2.6.1 On galaxy bias

In this study, we have taken a very simplified approach to galaxy bias, assuming that it is a constant over the redshift range of interest, and is not varied as a free parameter. In addition to assuming that the bias is well known and fixed, it is also the

same for both photometric and spectroscopic samples. These assumptions simplify the analysis and allow for the principle of redshift determination to be explored, but they are not realistic for a practical galaxy survey. Galaxy bias evolves in time (redshift) due to both observational biases (the selection function) and the evolution of the intrinsic galaxy bias.

The observational contribution to galaxy bias, through the identification of target galaxies, is unlikely to be independent of redshift. Since galaxies which are further away are fainter and redshifted then, in a magnitude limited survey, we will only see the brightest galaxies at high z. This population of galaxies might have very different biasing properties to the populations that we observe at low z, which should be a more complete sample.

The intrinsic galaxy bias is dependent on both the bias in galaxy formation, which will be based on the density distribution at the time when the galaxies form, and then the evolution of this bias as the galaxies move under gravity and the density fluctuations of matter continue to evolve. Many models such as the one proposed by Fry [158] and expanded by Clerkin et al. [43] suggest that galaxy bias should be inversely proportional to the growth. This is assuming that galaxies are formed at some early epoch and then move according to the cosmological gravitational potential, thereby relaxing the initial high bias and tending to increasingly trace the dark matter distribution. Other models which look at more complex features of the galaxy population such as merging [159], or those which calibrate to simulations [160], reach similar conclusions and propose a bias which increases with z, although the parameterisations and shapes of the bias functions vary.

Furthermore, photometric and spectroscopic surveys are well adapted to observing different populations of galaxies since they identify different features in their spectra. If these different populations have different bias properties, then the same bias parameters should not be applied to both photometric and spectroscopic galaxy samples. This can further complicate the use of, for example, the general time dependent (GTD) bias model, which parameterises the bias. One would be required to select a different set of bias parameters for different samples, increas-

ing the parameter space. This is also true for any photometric samples which are collected from different surveys and therefore optimised in different ways. This additional flexibility in the power spectrum, for both auto and cross correlations, will weaken our ability to constrain both cosmological parameters and the redshift distribution. In particular, we should be concerned with the difference between the spectroscopic and photometric sample biases, which affects the cross correlations.

The most important extension to this work will be to include a flexible galaxy bias prescription, and combine with weak lensing information to ascertain if the challenges in galaxy bias estimation can be overcome. Simplistically, for a given photometric sample p and spectroscopic sample s, the power spectrum of number counts depends on the biases

$$P(k)_{ps} \sim b_p b_s P(k)_{\delta\delta}, \tag{2.25}$$

$$P(k)_{pp} \sim b_p^2 P(k)_{\delta\delta}, \tag{2.26}$$

$$P(k)_{ss} \sim b_s^2 P(k)_{\delta\delta}, \tag{2.27}$$

where we have assumed for an individual bin we can approximate the bias as a constant. The bias of the photometric sample b_p can be constrained by comparison between the galaxy number counts and weak lensing in the photometric sample. Degeneracy between the redshift parameters and the bias is broken in this combination since the bias affects only the number counts, but the redshift distribution affects both (although in slightly different ways, as we shall see in the next chapter). Since the spectroscopic sample will likely be too small for a weak lensing analysis, we are left to infer b_s through the cross-correlation and through the height of the number counts spectrum. Since the redshift distribution of the spectroscopic sample is considered fixed however, the parameter b_s can potentially be quite strongly constrained, again without degeneracy with $n(z)$, and therefore $b_p b_s$, the bias of the cross-correlation, can be derived. It is therefore not unfeasible that there is enough information in a photometric and spectroscopic survey using both number counts and weak lensing to constrain both bias and $n(z)$ to some degree.

2.7 Further applications to non-Gaussian $n(z)$

The methods described in this chapter can be also be applied when $n(z)$ is not Gaussian for the photometric redshift bin(s) in question. In this case we need to explore the possible space of functions of redshift which are related to our initial estimate of $n(z)$. One possible way of doing this is to map the z coordinate to some new coordinate z'. This way, the transformations can be applied even if we do not have an analytic expression for $n(z)$, such as distributions estimated from real data-sets.

2.7.1 Details of n(z) Transformations

We will first describe in detail the transformations made to the $n(z)$ functions, and the motivations for the heuristic shape manipulation. It is not important that the higher moments are not exactly represented in the same way that the mean and variance of the gaussian are – these are after all only parameters controlling the shape which will be marginalised over. We are not really interested in what the skew or kurtosis of a bin actually is! The important thing is that it can explore a variety of shapes with a small number of parameters. More precise handling of the distributions can be achieved at the cost of increasing the parameter space, which may rapidly make the computation unmanageable without abundant computing resources.

2.7.1.1 Mean and Variance

Due to the spline representation used in our computation (see section 2.4.3) it is simplest to perform a transformation of the z axis to a new variable $z' = f(z)$ for a general (non-Gaussian) distribution. By creating the function $n'(z') = n(z) = n(f^{-1}(z))$, we obtain a distorted distribution in our new variable, which we take to be the new redshift. (This distribution, as all distributions, is normalised before any further calculations are carried out.)

To vary the mean ($\mu \to \mu + \Delta\mu$), we apply the transformation:

$$z' = z + \Delta\mu. \tag{2.28}$$

This varies the mean of an arbitrary distribution without affecting any of the higher

moments. We can also change the standard deviation ($\sigma \to \sigma + \Delta\sigma$) without affecting the mean or higher moments. (The fourth moment m_4 is changed, but not kurtosis $\kappa = \frac{m_4}{\sigma^4}$.)

$$z' = \left(1 + \frac{\Delta\sigma}{\sigma}\right)(z - \mu) + \mu. \tag{2.29}$$

One is free to change the mean and standard deviation of distributions in whichever order one desires as these transformations are commutative. Higher order transforms may also be applied which break symmetries, so care must be taken there as the operations will not commute.

2.7.1.2 Skewness

To adjust the apparent skewness, we need to stretch the distribution on one side of the mean, and squeeze the distribution on the other. For the sake of simplicity, we write a heuristic skew function controlled by a single parameter s. (This notation is to distinguish it from the genuine skewness calculated from the third moment, γ.) The parameter range is $-1 < s < 1$. We map from the original redshift coordinate z to a new coordinate z' representing the new redshift after the distortion has been taken into account. If the two are identical then we have $\frac{dz'}{dz} = 1$ everywhere. If we wish to stretch a region then $\frac{dz'}{dz} > 1$ and to squeeze it we have $\frac{dz'}{dz} < 1$. To achieve skewness, we need to smoothly vary from stretched regions on the one side of the mean, to squeezed regions on the other side, with $\frac{dz'}{dz}|_{z=\mu} = 1$.

We may choose a simple linear function:

$$\frac{dz'}{dz} = 1 + \frac{(z - \mu)s}{L}. \tag{2.30}$$

This fulfils the criteria discussed in the range $\mu - L < z < \mu + L$. After these points we fix $\frac{dz'}{dz}|_{z<\mu-L} = \frac{dz'}{dz}|_{z=\mu-L}$ and $\frac{dz'}{dz}|_{z>\mu+L} = \frac{dz'}{dz}|_{z=\mu+L}$. This is because otherwise we rapidly end up with very extreme stretching or squeezing of the distribution. Here we have an extra free parameter, L (the lengthscale of the skewness function). To avoid overburdening the routine with extra parameters we generically set this to $L = \frac{3\sigma}{4}$. The resulting function is found by integrating these expressions with the condition that $\mu' = \mu$. (This means that skewness does not interfere with

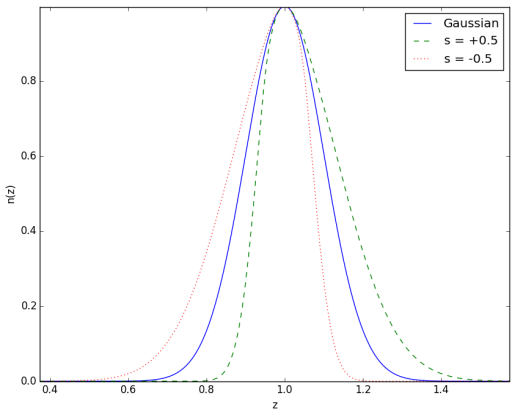


Figure 2.5: Results of applying positive and negative skew transformations to a gaussian photometric bin.

the peak of the distribution, but will change the mean; it may also interfere with standard deviation. Standard deviation and mean can separately be readjusted to remove this degeneracy if desired.)

$$z' = \begin{cases} \frac{s}{L}\left(\frac{1}{2}(z+\mu)^2 - \mu\right) + z, & |z-\mu| \leq L \\ (\mu+L)' + (1+s)(z-(\mu+L)), & z-\mu > L \\ (\mu-L)' + (1-s)(z-(\mu-L)), & z-\mu < -L. \end{cases} \tag{2.31}$$

2.7.1.3 Kurtosis

Kurtosis is handled in a similar way to skewness, by a heuristic function controlled by a single parameter k (different to the exact kurtosis, traditionally κ) which varies between $-1 < k < 1$. This case is symmetric around the mean, and we wish to stretch the distribution close to the mean, and squeeze it further away (or vice versa).

In this case we again need to choose length scales. We choose to have the transition from stretched to squeezed regions (i.e. $\frac{dz'}{dz} = 1$) at $(z-\mu) = \sigma$, and then fix the transformations (as with the skewness above) at $(z-\mu) = 2\sigma$. This requires in the below expression $L = \sigma$.

For simplicitly we again choose linear relations.

$$\frac{dz'}{dz} = (1+k) - \frac{|z-\mu|}{L}k. \tag{2.32}$$

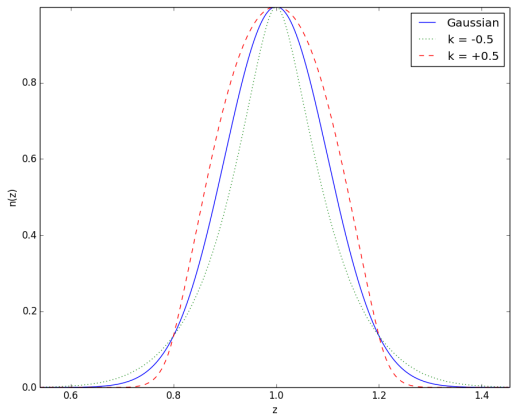


Figure 2.6: Results of applying positive and negative kurtosis transformations to a gaussian photometric bin.

We then integrate as before, choosing $\mu' = \mu$. Once again, the standard deviation can be separately adjusted for if desired.

$$z' = \begin{cases} z + k(z-\mu) - k\frac{(z-\mu)^2}{2\sigma}, & \mu \leq z \leq \mu + 2\sigma \\ z + k(z-\mu) + k\frac{(z-\mu)^2}{2\sigma}, & \mu - 2\sigma \leq z \leq \mu \\ \mu + \sigma + (1-k)(z-\mu-2\sigma), & z \geq \mu + 2\sigma \\ \mu - \sigma + (1-k)(z-\mu+2\sigma), & z \leq \mu - 2\sigma. \end{cases} \tag{2.33}$$

2.7.2 Constraining higher moments

As a benchmark, we also present constraints on the shapes of a single photometric bin from cross correlations with three spectroscopic bins. Here we have a fixed cosmology to simplify the calculations and provide benchmark results for the shape parameters. (This is obviously not realistic, although we have seen in section 2.5.3 and fig 2.4 that the degeneracies with cosmology are minimal when μ and σ are well constrained by spectroscopy.) With cross correlations between more photometric and spectroscopic bins we expect these results to be improved.

We find that the odd moments are strongly correlated, but we also have less information to constrain the higher moments from the correlation statistics. This may not be a problem unless the higher moments significantly affect the cosmo-

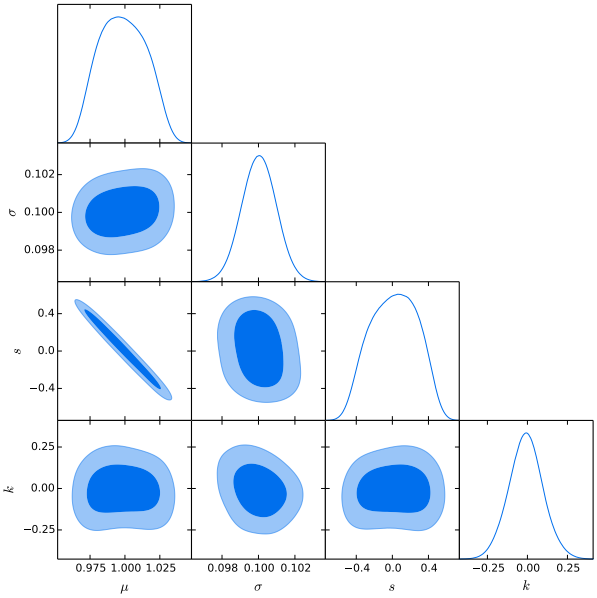


Figure 2.7: Constraints on the mean, width, and shapes of bins as characterised by (μ, σ, s, k) for a single photometric bin, cross correlated with three spectroscopic bins, with fixed cosmology.

logical parameter estimation, although the lack of constraining power of the $C(l)$s suggest that these are not strongly affected by the finer details of the shape of the distributions. The most significant potential problem is the degeneracy between s and μ. This could lead to a spreading of the distribution over μ which may affect the cosmological parameter estimation. In order to combat this one would need to have relatively dense spectroscopic samples throughout the range. In principle, the degeneracy between μ and s need not be a problem. Since they are strongly degenerate, it is largely the case of the one compensating for the effect of the other. In this case, the cosmological calculation may not be strongly affected even if the uncertainty in μ increases significantly. We should also bear in mind that the application of the skew transformation alters the mean of the distribution even though the peak is kept in the same place. This needs to be compensated for by the μ parameter, and thus the true mean may remain roughly unchanged and be very strongly constrained. We expect higher order moments to have less of an impact on results that μ or σ, but nevertheless may be important particularly if the shape of the bin is highly uncertain. A redefinition of the skew transformation to account for the change in mean should de-correlate μ and s, and then the shapes should be independently

constrainable, provided that one can handle the large increase in parameter space.

2.7.3 Identifying anomalous features in $n(z)$

Another example, closely related to the skew and kurtosis of a bin, is the possibility of a tail on one or both sides of the distribution, which is previously unknown. How might such an artefact show up in the $C(l)$s, and can we catch and correct such problems? A redshift tail may affect the correlations in a few different ways. Firstly, having a tail spreads the $n(z)$ distribution over a larger space. This lowers the peak (by normalisation), and reduces correlations in the sample (since galaxies which are widely spaced in redshift are less correlated). Since these affect predominantly the amplitude of the autocorrelation, this is highly degenerate with cosmology. The cross-correlations on the other hand can reveal the redshift distributions; a bin with a high redshift tail will show strong correlations with redshift bins significantly higher than the mean, but not with those significantly lower than the mean. This cannot be replicated using a bin without tails (e.g. Gaussian, top-hat, or so on) and biased cosmological parameters, since cosmological parameters can't compensate to boost the correlations in specific bins, but affect all the correlations. Such signals in the data are red flags that the $n(z)$ modelling is non-trivial and should be investigated further.

For this example we will take two bins, one with a high redshift tail, and one with a bump, and look at their cross-spectra with a spectroscopic redshift sample covering the redshift range. The redshift distributions are shown in fig 2.8. The cross-correlations are calculated with a series of spectroscopic bins of width $\sigma = 0.03$ and means evenly spaced throughout the redshift range, $\mu = \{0.3, 0.6, 0.9, 1.2, 1.5, 1.8\}$. This should be enough to illustrate the effect that would be had in a dense spectroscopic sampling. The correlations are plotted in fig 2.9; fig 2.9a shows the cross-correlations for Bin 0, which has a high redshift tail, whilst fig 2.9b shows the cross-correlations for Bin 1, which has a high redshift bump. We can see the tell-tale signs of these anomalous features in the $C(l)$s, and will know that these cannot be produced by Gaussian bins.

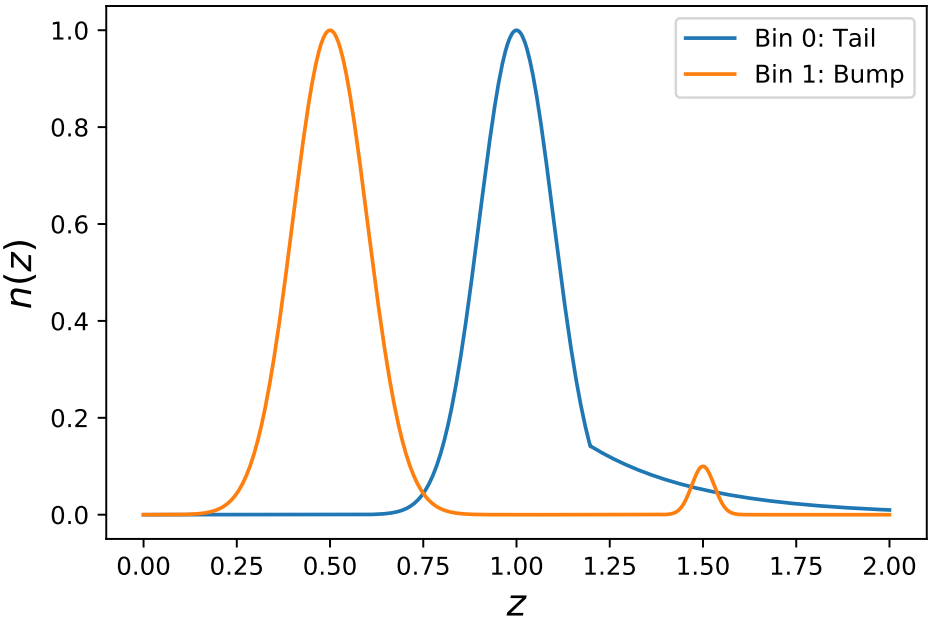
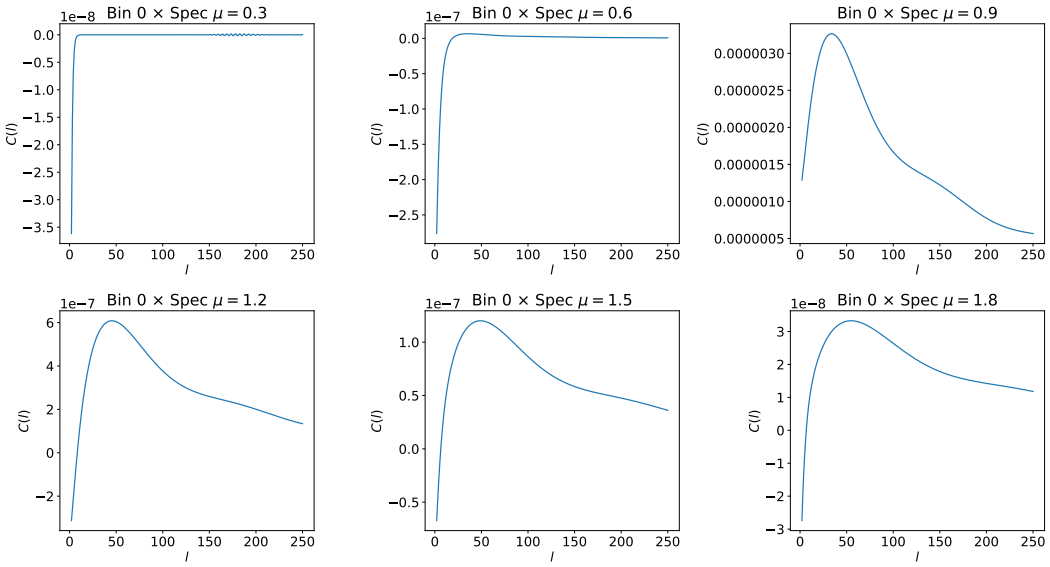
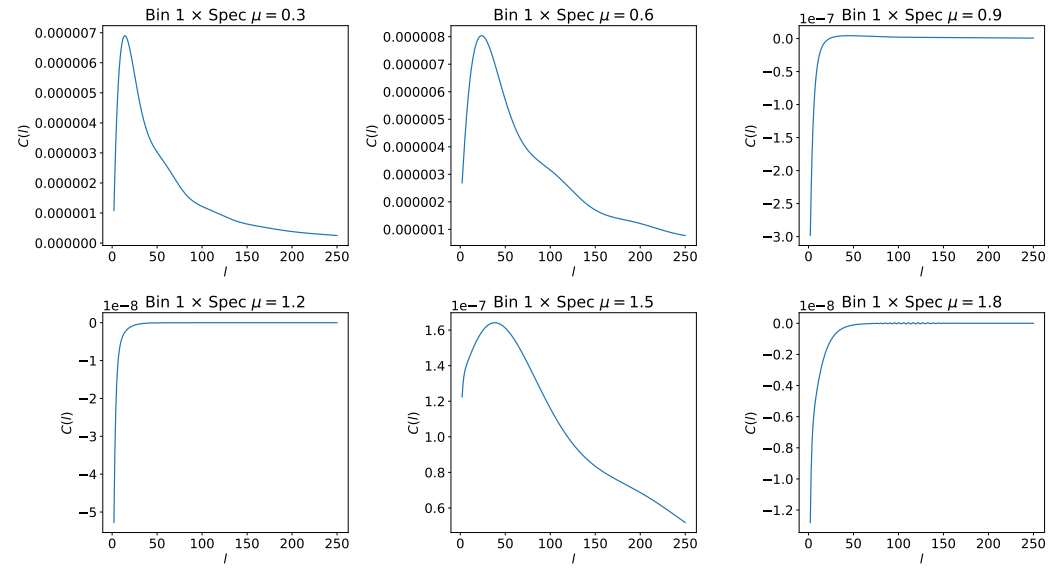


Figure 2.8: Correlations functions for bins with a high redshift tail (a), and a high redshift bump (b) in $n(z)$. These features can be clearly seen in the cross-spectra by noting the asymmetry of the cross-correlations around the mean of Bin 0 ($\mu = 0.1$), and by noticing the boost in correlation with an isolated spectroscopic sample with Bin 1 (corresponding to the bump).



(a)



(b)

Figure 2.9: Cross-correlation functions for six spectroscopic bins with photometric bins with a high redshift tail (a), and a high redshift bump (b) in $n(z)$. These features can be clearly seen in the cross-spectra by noting the asymmetry of the cross-correlations around the mean of Bin 0 ($\mu = 0.1$), and by noticing the boost in correlation with an isolated spectroscopic sample with Bin 1 (corresponding to the bump).

Chapter 3

Horndeski Gravity using Large Scale Structure and Weak Lensing

Merely cosmological detail
intended to lend artistic
verisimilitude to an otherwise bald
and unconvincing narrative.

Pooh-Bah, Lord High Astronomer
of Titipu

3.1 Introduction

We continue along similar lines to the previous chapter, but we will extend our approach to include a more realistic survey setup including a large number of spectroscopic bins spread over a wide range. We also explore physics beyond ΛCDM by performing a modified gravity analysis using the HICLASS code [132] to calculate the effects of phenomenological scalar field models. The nature of dark energy (DE), the source of the universe's late time acceleration, is one of the most pressing concerns in modern cosmology. So much so that over four hundred scientists are currently working on the 'Dark Energy Survery' (even if not all of them are working on DE) [139]. Its initial and simplest attribution to a cosmological constant (Λ) has, with good reason, held its place as the most popular option. It fits the data very nicely, and is by far the simplest way to resolve our observed late time expansion. It

requires no additions to GR which do not appear straightforwardly in its derivation (Λ having been previously invoked by Einstein for the purpose of creating a static universe model). Whilst a cosmological constant provides an answer, it's not the only option, and there are good reasons to explore the sensitivity of observables to larger changes in the theory. Nevertheless, the effects of dark energy must (by virtue of observational evidence so far) be strikingly similar to a cosmological constant. The evolution of the background is tightly constrained by a variety of probes including supernovae, large scale structure, and the CMB; a great deal of focus has been turned instead to finding non-Λ effects in the perturbations instead. Whilst there may be some hope of seeing non-GR effects or additional dynamical energy components in DES, most people are pinning their hopes on the next generation of experiments. As an aside, almost all dark energy or modified gravity models contain Λ as a limit point, so if we really do have dark energy that is indistinguishable from Λ, then all that it is possible to do is squeeze the error bars, but never eliminate the models entirely. When working with models which are not necessarily falsifiable, it is an interesting question to consider when one ought to give up; nevertheless a significant number of people – myself included – are saying 'not yet! (But maybe soon.)' With Euclid ([142][143][144]) and LSST ([145][146][147]) providing a wealth of new information on large scale structure, and gravitational wave experiments already bearing some of the most drastic results in the history of the field, the next generation of experiments will be a crucial hurdle for the survival of such theories.

The biggest problem with the 'dark sector' of physics is the overabundance of possible models which can explain what we see. Much like the situation with dark matter, scalar-tensor theories (and DE/MG models in general) have proliferated during their short lifetime and there are stacks of papers outlining different possible models, many of which are variants on the same basic themes. An overarching framework can be worked out to classify theories, but since the framework involves free functions each choice of function can produce a different theory with different predictions. Some may be unstable and unable to produce reliable predictions of

any kind. When approaching MG with a view to observable results, one typically chooses a phenomenological parameterisation which captures the effects of the dark energy, without a fundamental theory. Rather than using general functions, the time evolution is usually fixed and related in some way to the fractional density of dark energy $\Omega_{\rm DE}(z)$. This allows one to search for departures from GR rather than try to detect individual models, and allows one to constrain directly the effects on growth and lensing. On the other hand, it gives no indication of how to construct a theory at the level of the action which will reproduce these effects, nor does it tell us anything about the nature of the DE field except that it is not a constant. Crucially, results may also be strongly dependent on the choice of time evolution, and searching for deviations with the wrong time dependence could conceivably lead to a failure to make a detection. Whilst parameterisations have been considered which address the lack of correspondence to the action, as yet we are at a loss to explore the infinitely various possible time dependencies in a systematic and computationally tractable way.

When investigating the growth of structure in modified gravity models in some parameterisation, one typically fixes the background to ΛCDM or wCDM dynamics. The underlying theory, if reconstructed, must be able to reproduce this limit without sacrificing its effects on the perturbations. The potentials in many scalar-tensor theories are able to take the place of Λ, often via a constant V_0, although this is equivalent to putting Λ back in! The question is not necessarily 'is there a scalar field instead of Λ?', but rather 'is there a scalar field?'

The perturbations are analysed through the CMB and galaxy surveys, which require vast amounts of data in order to produce precise enough power spectra to beat down uncertainties and constrain the growth of structure in the universe. The Planck mission [76] has produced the finest CMB dataset to date, combining temperature, polarisation, and lensing information into a powerful cosmological probe. Modified gravity analyses benefit from a multi-probe approach however, and future galaxy surveys will provide excellent complementary datasets. The Euclid mission is a satellite based galaxy survey which is planned for launch in 2020. It has the

potential to vastly improve our understanding of the universe on large scales, by imaging an estimated 1.5 billion galaxies for photometry and shape measurement. These galaxies can be used to constrain cosmology by calculating the power spectrum of number counts and weak lensing shear; the key here is that these observables probe complementary aspects of the physics. One tracks the motion of matter through space by measuring the gravitational growth of perturbations, and the other the motion of photons by measuring the deflection of light by mass. This provides information about both Bardeen potentials, Φ and Ψ, which encode the fluctuations in the space time itself and provide our most direct probe of gravity. A combination of photometric and spectroscopic data will be present in unprecedented volumes, and joint analyses with CMB information (Planck) will can break degeneracies and constrain parameters which have little effect on the growth of structure.

We examine the prospects of Euclid, combined with CMB temperature information, for constraining Horndeski models of gravity. As in the previous chapter, the analysis will jointly constrain the photometric redshift distribution, to examine the prospects of using this technique on larger numbers of bins and outside of standard ΛCDM. This will help us to know if our photometric redshift uncertainties are liable to influence the results of MG analyses in Euclid.

3.2 Scalar-Tensor Modified Gravity and Parameterisation

Scalar-tensor theories are variations of the GR action which include a scalar field that couples to the curvature (or to the matter action via a conformal transformation). The dynamics of scalar-tensor theories may be extremely complicated, especially on small scales. In order to pass the stringent gravitational tests which have been performed using the solar system [154], gravity on small scales or in high density environments must be 'screened' in some way [67][68]. This usually involves the scalar field vanishing in high density environments, or its effective mass becoming very large (and thus its effective force short ranged) so that it cannot impact the dynamics. This level of detail is impractical in cosmology, and the nonlinear scales

in modified gravity are not well understood. The perturbation equations may be calculated, but these may also be extremely complicated in the general case. They will in general depend on the evolution of the scalar field $\varphi(k,\chi)$ and its kinetic term; the scalar field rarely has a simple equation of state. Furthermore if one were to model the evolution of the scalar field and its impact on dynamics explicitly, one would still have an infinite number of theories to contend with. In order to proceed it makes more sense to parameterise the functions in some way which will make the calculations more tractable, and avoid having to explore theories on an individual basis.

For LSS, these theories are often cast as effective field theories (EFT), which allow us to break down the possible behaviour of the action in to manageable chunks. The impact on perturbations can be reduced to a small collection of free functions, which can then be parameterised directly in the analysis. Reconstruction of the fundamental theory would require understanding how the form of the Lagrangian is related to the form of the perturbing functions.

3.2.1 The Horndeski action

The Horndeski formalism rewrites such Lagrangians as a sum of Lagrangian contributions, $\mathscr{L}^m$ (for the matter sector), and four contributions $\mathscr{L}^{2-5}$ for the gravity/scalar field sector.

$$S = \int \sqrt{-g}\left[\sum_{i=2}^{5}\mathscr{L}_i + \mathscr{L}_m[g_{\mu\nu}]\right], \tag{3.1}$$

$$\mathscr{L}_2 = K, \tag{3.2}$$

$$\mathscr{L}_3 = -G_3\Box\varphi, \tag{3.3}$$

$$\mathscr{L}_4 = G_4 R + \frac{dG_4}{dX}\left[(\Box\varphi)^2 - \nabla_\mu\nabla_\nu\varphi\nabla^\mu\nabla^\nu\varphi\right], \tag{3.4}$$

$$\mathscr{L}_5 = G_5 G_{\mu\nu}\nabla^\mu\nabla^\nu\varphi \tag{3.5}$$

$$-\frac{1}{6}\frac{dG_5}{dX}\left[(\Box\varphi)^3 + 2(\nabla_\mu\nabla^\nu\varphi)(\nabla_\nu\nabla^\alpha\varphi)(\nabla_\alpha\nabla^\mu\varphi) - 3\nabla_\mu\nabla_\nu\varphi\nabla^\mu\nabla^\nu\varphi\Box\varphi\right], \tag{3.6}$$

where K and G_i are free functions of ϕ and $X = \frac{(\nabla\phi)^2}{2}$. In fact, it was originally derived with no matter Lagrangian (or mention of the word gravity at all), but rather exploring possible actions of a scalar field interacting with a metric (the 'tensor' part of scalar-tensor). It is the most general scalar-tensor action which can be written such that the following desirable properties are satisfied:

- Only the metric, $g_{\mu\nu}$, and a single scalar field, ϕ, are included.
- The action is at most second order in derivatives of $g_{\mu\nu}$ and ϕ.
- The resulting equations of motion are at most second order in derivatives of $g_{\mu\nu}$ and ϕ.

The original form of the action was derived by Horndeski in [133]. For the purposes of cosmology, it is typically rewritten in the form we will give here, which appears in most of the contemporary literature such as in [153].

This gives us all of the terms that we are allowed to use in a so called 'Horndeski' theory of gravity. It is plain to see that the majority of scalar-tensor theories in the literature are included in this class, with simple theories leaving many of the terms out completely.

3.2.2 The α-functions and parameterisation

Closely related to other effective field theory (EFT) parameterisations, Bellini et. al. express the perturbations in Horndeski theories in terms of four free functions: α_M, α_K, α_B, and α_T [153]. The benefit of this parameterisation is that these have,

to an extent, a physical interpretation that also relates relatively straightforwardly to the action itself. Furthermore, it avoids redundancies in the description by combining EFT functions which only appear in certain combinations in the perturbation equations. Lastly, there has been developed a version of the CLASS code called HICLASS [132] ('Horndeski in CLASS') which may be used straightforwardly as a Boltzmann engine for this parameterisation. The α-functions and the effective Planck mass M_*^2 can be defined in terms of the Lagrangian (see the appendix of [153] or section 1.110 of this thesis for the expressions). As a result, it is straightforward to calculate the contributions to each α-function from an arbitrary scalar field action. Each of the functions in this parameterisation has a different role in the dynamics.

- M_*: The 'effective Planck mass'. Another way to think of this (which is perhaps more intuitive in the context of a classical field theory and for comparisons with GR and Newtontian gravitation) is the effective gravitational constant $G_N \to G_*(z)$. This controls the strength of gravity on cosmological scales; if it is different from the GR theory then there must be screening on small scales to evade solar system constraints. The parameterisations are not detailed enough to investigate the properties of screening, so one must assume that one can construct a theory with an appropriate mechanism, of which there are many examples in the literature (chameleon, Vainstein etc.).

- α_M: The 'running of the Planck Mass'. This encodes the variation of M_* (or, equivalently, G_*) over time: $\alpha_M = \frac{1}{H}\frac{d\ln M_*^2}{dt}$. If this parameter is used, then M_* is simply set at the present epoch, and then inferred for the rest of time from α_M.

- α_K: The 'kineticity' of the dark energy. This term has minimal effects on large scale structure except at the very largest scales [132] (where it is often a struggle to get good data). It is the most fundamental parameter in the sense that it has contributions from each part of the Lagrangian, and is thus present in all theories even down to a perfect fluid; if the other α-terms are zero there

are no couplings between the scalar field and the geometry. Then we are left with a minimally coupled scalar field which only affects the background, and hence in the parameterised model with a fixed expansion history it can have no effect. (The effects of perfect fluids on the background are modelled with an effective equation of state w.) When the other terms are present it suppresses the sound speed of dark energy, and leads to a boost in structure at very small k. (It can noticeably affect larger scales if the value is very large.)

- α_B: The 'braiding' term. This term is generated by parts of the Lagrangian that couple the kinetic term $X = \frac{1}{2}(\nabla\phi)^2$ with the geometry via $g^{\mu\nu}$, R, or $G_{\mu\nu}$. If one eliminates α_T in the simplest way (without algebraic cancelation), then X cannot couple directly to R or $G_{\mu\nu}$, and couples only to geometry via $g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$.

- α_T: The 'tensor excess'. This term determines the difference between the speed of gravitational waves and light, $\frac{v_{\rm GW}}{c} - 1$. Its contributions from the Lagrangian are quite limited, and theories with no tensor excess $\alpha_T \equiv 0$ are straightforward to construct. These include most basic scalar-tensor theories in the literature.

The parameter α_T has been the subject of great attention lately. For the first time, gravitational waves have been spotted with an electromagnetic counterpart in the event GW170817 [150]. The fact that the gravitational waves and the electromagnetic signal we observed so close together in time suggests that the speed of gravitational waves must be very close to the speed of light. Studies such as [148] have shown that $\alpha \lesssim 10^{-15}$ at the present epoch. This places extremely tight constraints on α_T, especially in most phenomenological models in which the parameterised functions are monotonically increasing. Section 1.2.3.1 contains a fuller description of the actions now permissible, but for the purposes of this chapter we can simply remove α_T from the analysis. We perform therefore two version of our forecast: one with α_T free to vary, and one with $\alpha_T = 0$.

Parameterisation of the evolution of the α-functions presents many options.

The most common modified gravity parameterisation is to set modifications proportional to $\Omega_{\rm DE}$. This is also a common parameterisation in the Horndeski case,

$$\alpha_i(z) = \alpha_i(0) \cdot \frac{\Omega_{\rm DE}(z)}{\Omega_{\rm DE}(0)}. \tag{3.7}$$

This is the approach taken in [138] and [137]. In [137] it is argued that many MG theories ought to be able to be described this way, and a prescription for calculating how to do this from the action is suggested.

Notice that there is a significant parameter space in this parameterisation which leads to instabilities. A significant cause of instability, particularly with respect to the parameters M_*^2 and α_M, is the so called 'gradient instability' where $c_s^2 < 0$. To understand why these systems are so prone to these instabilities, we can look at the expression for the scalar sound speed c_s^2:

$$c_s^2 = \frac{-(2-\alpha_B)\left[\dot{H} - H^2(\alpha_M + \frac{1}{2}\alpha_B)\right] - H\dot{\alpha}_B + \frac{8\pi G_N}{M_*^2}\rho_m}{H^2(\alpha_K + \frac{3}{2}\alpha_B^2)} > 0, \tag{3.8}$$

where in this expression we have simplified by setting $\alpha_T = 0$, assumed that the matter is pressureless, and made the dependence on M_*^2 explicit. Stability requires that the denominator be always positive definite, so we will only look at the numerator. As an example, let us consider a case where we have a cosmological strength of gravity which is different from that in the solar system i.e. $M_* \neq 1$. Let's also assume for now that we have no other modifications. Then we have a simple inequality

$$-2\dot{H} - \frac{8\pi G_N \rho_m}{M_*^2} > 0. \tag{3.9}$$

We may now compare this to the expression for H^2 in ΛCDM (to which our models will tend as $\alpha_i \to 0$ and $M_* \to 1$),

$$H^2 = H_0^2 \sum_i \Omega_i a^{-3(1+w_i)}. \tag{3.10}$$

Considering for now just a case with matter and dark energy, we have upon differ-

entiation and rearrangement,

$$2\dot{H} + 8\pi G_N \rho_m = 0, \tag{3.11}$$

we see immediately that if $M_*^2 < 1$ then the inequality 3.9 is violated. These kinds of instabilities can also occur when $\alpha_M \neq 0$; the braiding (or tensor excess) can have a stabilising effect but nevertheless these parameters are greatly constrained by stability concerns, and $M_*^2 < 1$ will almost certainly enter an unstable regime at some point. (Since the parameterisation of the α-functions is proportional to the fractional density of DE, in the early universe $\alpha_i \to 0$ and the stability criterion tends towards equation 3.9.) It should be noted that if $\alpha_B = 0$, then the only contribution to M_*^2 is from a constant coupling to the curvature aR i.e. a redefinition of Newton's constant. Nevertheless, it is a useful example to show that when the α-functions become very small and the solution becomes extremely close to ΛCDM, small deviations can lead to instabilities.

Perhaps even more pressing is the question of whether or not any of the popular parameterisations are appropriate at all. They are not generally motivated by strong theoretical concerns, nor do any attempt to cover the full space of possible functions (which would likely be computationally intractable). One might ask how strongly our ability to detect modified gravity effects depends on our choosing a parameterisation which comes close enough to the correct scalar field evolution! Although most scalar-tensor theories explored in the literature (comprising the bulk of cosmological modified gravity models) come under the umbrella of the Horndeski class, they each have their own specific time evolutions. $f(R)$ gravity, for example, has $\alpha_K = \alpha_T = 0$ and $\alpha_M = -\alpha_B$, but the time evolution of $\alpha_{M,B}$ may be very different from the parameterisation that we have considered. Although it is beyond the scope of this chapter, in future work it would be interesting to calculate the perturbations for some well known specific models, and see if these lead to detections of non-GR when analysed in these common parameterised frameworks, and how close the Lagrangian contributions to the best fit model come to resembling the action of the fiducial theory under consideration.

3.2.3 Modified Growth and Lensing in Horndeski Gravity

Deviations from GR in modified gravity are often described in terms of two variables $\mu(k,\chi)$ and $\Sigma(k,\chi)$ [46][141]. These represent changes to the linearised Poisson equations for matter perturbations:

$$-k^2\Psi(k,\chi) = 4\pi G\bar{\rho}(\chi)\delta(k,\chi)\cdot\mu(k,\chi), \tag{3.12}$$

$$-k^2\left[\Phi(k,\chi)+\Psi(k,\chi)\right] = 8\pi G\bar{\rho}(\chi)\delta(k,\chi)\cdot\Sigma(k,\chi). \tag{3.13}$$

In the ΛCDM+GR concordance model $\Phi=\Psi$ and $\mu=\Sigma=1$. In the late time regime which we are looking at $z\lesssim 2$, we expect that we may ignore the contributions to these equations from relativistic species such as radiation or neutrinos. We may also calculate the gravitational slip parameter

$$\eta = \frac{2\Psi}{\Phi+\Psi}. \tag{3.14}$$

In the usual ΛCDM formulation (with no torsion etc.) the parameter $\eta=1$ since $\Psi=\Phi$.

These functions are the ones typically directly parameterised by phenomenological approaches to modified gravity. Although in the approach we have adopted these functions are not directly parameterised, they can be inferred by comparison of the matter perturbations to the Bardeen potentials, which will reveal the effective μ and Σ functions as their ratio. This allows us to grasp intuitively the effects which the modified gravity model are having, and also compare the results to approaches which model deviations at the level of the equations above. A thorough discussion in the 'extreme quasi-static limit' (quasi-static limit with $k\to\infty$) can be found in [153], as well as an analytic conversion from the α-functions to μ, Σ, and η in this limit.

3.2.3.1 Example 1: Modifying the Growth through M_* or G_*

In the case of a varying gravitational strength $G \to G_*(\chi)$, we could express this as

$$G_*(\chi) = G_N \mu(\chi) \implies \mu(\chi) = \frac{G_*}{G_N}, \tag{3.15}$$

or

$$\mu(\chi) = \frac{1}{M_*^2}, \tag{3.16}$$

and thus the modified growth $\mu(\chi)$ would contain the effects of new strength of gravity on large scales. We can see that if we increase M_* (and therefore lower G_*), then the growth modification $\mu(\chi)$ drops accordingly, damping the growth of large scale structure. The lensing term Σ is affected likewise, but this will introduce no gravitational slip (as all we are doing is changing a fundamental constant). A model such as this is indistinguishable from ΛCDM, because it is identical to a ΛCDM model with an appropriately adjusted matter component $\rho_{\Lambda\text{CDM}} = \rho_{\text{MG}} M_*^2$.

When $\alpha_M \neq 0$, then the effective gravitational constant can change, and this is no longer precisely degenerate with adjusting the mass, and therefore is distinguishable from ΛCDM. Fig 3.1 shows our modification functions in the case where $M_{p,0}^2 = 1.05$, and $\alpha_M = 0.5$ i.e. the effective Planck mass increases over time. In this case, at high redshift (where Ω_{DE} is subdominant and thus $\alpha_M(z)$ is small), we have a case where $\mu = \Sigma = M_{p,0}^{-2}$ as one would expect for a constant change in M_*^2 (and thus $\eta = 1$). As α_M grows, we see the growth increase and the potentials Ψ and Φ split to create gravitational slip. This means that lensing combined with galaxy counts should be an excellent probe for this sort of model. You may notice that even though α_M is positive, and therefore M_*^2 is increasing with time, the growth modification μ actually increases up until very late times. This is because α_M also enters into the perturbation equations in a non-trivial way. Looking at the extreme quasi-static limit for a model with only α_M and α_K (see [153]), we have that

$$\eta = 1 + \frac{2\alpha_M^2}{2c_s^2\alpha_K + 2\alpha_M^2}, \tag{3.17}$$

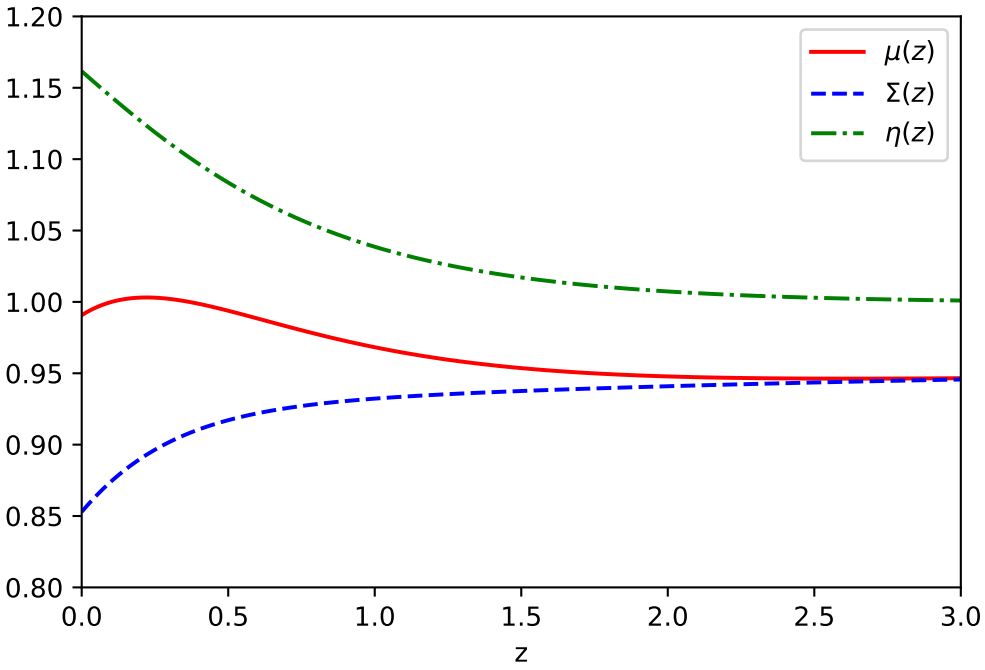


Figure 3.1: Modified growth (μ), lensing (Σ) and gravitational slip (η) parameters for a Horndeski model with $\alpha_M = 0.5$, $M^2_{*,0} = 1.05$.

and

$$\mu \approx M_*^{-2} \frac{\eta}{2-\eta}, \tag{3.18}$$

where we have made the dependence of μ on M_*^2 explicit. From equation 3.17 we have $\eta > 1$, thus the value of μ will depend on the balance between M_*^2 and η. Equation 3.17 suggests that $\eta \approx 1$ when α_M is small (i.e. early times) and thus $\mu \approx M_*^{-2}$. As dark energy begins to dominate, η becomes more significant, and we have overall growth in μ despite the increase in M_*^2. This is a straightforward example of how even a seemingly simple parameter in Horndeski gravity can actually affect the dynamics in unintuitive ways.

3.2.3.2 Example 2: Introducing Scale Dependence

The example we looked at above has not featured any scale dependence. Some Horndeski models do however exhibit some scale dependence in the functions μ and Σ. We can see this when we look at a model with $\alpha_M = -0.1$, $\alpha_B = 0.5$. (This time, we will keep $M^2_{p,0} = 1$.) The functions from this model are plotted in Fig 3.2;

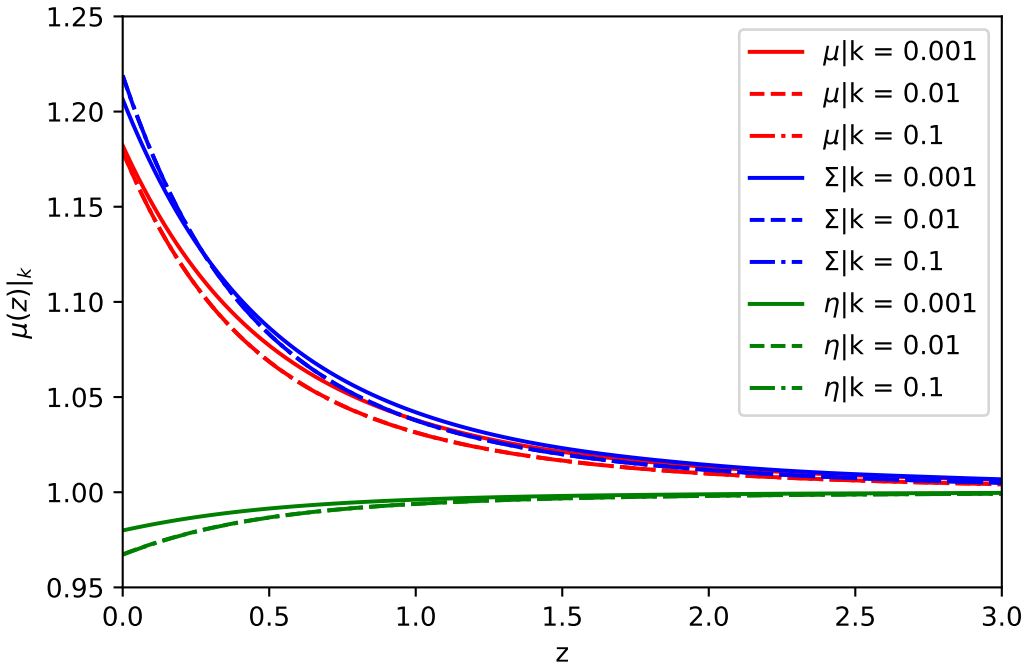


Figure 3.2: Modified growth (μ), lensing (Σ) and gravitational slip (η) parameters for a Horndeski model with $\alpha_M = -0.1$, $\alpha_B = 0.5$. Different line-styles indicate different values of k.

one can see that there is a small scale dependence at the largest scales (the $k = 0.01$ and $k = 0.1$ scales are indistinguishable).

3.3 $C(l)$s and Window Functions

The $C(l)$ formalism provides a powerful and flexible framework for calculating auto and cross correlation functions for multiple cosmological probes, which is vital for constraining modified gravity.

$$C_{ab}^{ij} = \frac{2}{\pi} \int W_a^i(l,k) W_b^j(l,k) P(k) k^2 \, dk \,. \tag{3.19}$$

Once we have calculated the window functions $W_a^i(l,k)$ for each of our probes and redshift bins, we may cross-correlate any combinations of them. In this chapter, we will take further advantage of this by introducing cosmic shear window functions and their cross correlations with galaxy number counts.

3.3.1 Galaxy Number Counts

The galaxy number count window function is given by

$$W_g^i(l,k) = \int_{\chi_{min}}^{\chi_{max}} b(\chi) n^i(\chi) D(k,\chi) j_l(k\chi)\, d\chi\,. \tag{3.20}$$

Note that, in comparison to equation 2.5 we now include the possibility of k-dependence in the growth function. In ΛCDM this is usually only necessary in the non-linear regime, but even the linear dynamics may be scale dependent in MG theories.

We also include redshift space distortion (RSD) as before,

$$W^i_{\rm RSD}(l,k) = f\int n(\chi)D(k,\chi)[\frac{2l^2+2l-1}{(2l+3)(2l-1)} j_l(k\chi) - \frac{l(l-1)}{(2l-1)(2l+1)} j_{l-2}(k\chi) - \frac{(l+1)(l+2)}{(2l+1)(2l+3)} j_{l+2}(k\chi)]d\chi, \tag{3.21}$$

where $f = \frac{d\log D}{d\log a}$, which is approximated as constant over the width of the bin. Redshift space distortions are sensitive to modified gravity through γ or $f(k,\chi) = \frac{d\log D}{d\log a}$, both of which measure the modified growth rate [141].

3.3.2 Weak Lensing Shear

Weak lensing requires a little extra care in MG/DE theories. In addition to being sourced by the modified growth of matter perturbations, it also responds to potential anisotropy in the metric, through a modified Poisson equation

$$k^2[\Phi(k,\chi)+\Psi(k,\chi)] = 8\pi G_N a^2 \rho\delta\cdot\Sigma(k,\chi). \tag{3.22}$$

This multiplicative factor of Σ enters into the lensing equation, since light travelling along null geodesics will feel the effective potential $[\Phi+\Psi]$. (It is the potential in which we are genuinely interested, so we merely use the matter perturbations as a proxy; expressing the equation in this forms makes it easier to compare with other parameterisations.) The form of $\Sigma(\chi)$ is given by the dynamics of the particular modified gravity theory, some examples of which we have shown in section 3.2 and

which are discussed in greater depth in [153].

The shear window function W_γ^i is described by a slightly more complex nested integration [46].

$$W_\gamma^i(l,k) = \int \Sigma(k,\chi) D(k,\chi) q(\chi) j_l(k\chi)\, d\chi, \tag{3.23}$$

where $q(\chi)$ is the lensing weight function. It is given by an integral over χ [31],

$$q(\chi) = \frac{3H_0^2\Omega_m}{2c^2}\frac{\chi}{a(\chi)}\int_{\chi_H}^{\chi} n(\chi')\left(1-\frac{\chi}{\chi'}\right)d\chi'. \tag{3.24}$$

Although we have cast the integrals in terms of χ rather than z, when we describe a bin as 'Gaussian', we mean that the distribution is Gaussian in z, and therefore somewhat skewed in χ. Nevertheless a simple transformation exists between them:

$$\frac{d\chi}{dz} = D_H E^{-1}(z) \implies n(\chi) = n(z)\frac{E(z)}{D_H}, \tag{3.25}$$

where $D_H = \frac{c}{H_0}$ is the Hubble distance.

3.4 The Likelihood Function

The likelihood function is the same as in section 2.4, extended to include the shear $C(l)$s and the cross correlations between probes,

$$\begin{aligned}\ln(L) = &\frac{f_s}{2}\sum_{l}^{l_{max}^{LSS}}(2l+1)\left[\mathrm{Tr}\left(I - M_{A,l}M_{B,l}^{-1}\right) + \ln\left(\det\left(M_{A,l}M_{B,l}^{-1}\right)\right)\right] \\ &+ \frac{f_s}{2}\sum_{l=2}^{l_{max}^{CMB}}(2l+1)\left[1 - \frac{C_{TT}^A(l)+N(l)}{C_{TT}^B(l)+N(l)} + \log\left(\frac{C_{TT}^A(l)+N(l)}{C_{TT}^B(l)+N(l)}\right)\right].\end{aligned} \tag{3.26}$$

Here we have ignored any cross correlation between LSS and the CMB. The LSS $C(l)$ matrices $M(l)$ are defined as

$$M(l) = \begin{pmatrix} C_{nn}(l) & C_{n\gamma}(l) \\ C_{\gamma n}(l) & C_{\gamma\gamma}(l)\end{pmatrix}, \tag{3.27}$$

which is symmetric and positive semi-definite by the symmetry properties of the C_{ab}^{ij} sub-matrices. In order for the likelihood to be calculable for the above procedure, the matrices M_B and $M_A M_B^{-1}$ must be non-degenerate.

We restrict l to the linear regime, as non-linear dynamics are not well worked out in MG theories (and even in ΛCDM the modelling is often too dubious and the uncertainties too large to confidently compare to data) and we are not using non-linear estimators. We use a prescription described in [152] to approximately identify the appropriate l-range for linear scale physics,

$$l_{\max} = k_{\max}\chi(\bar{z}) \sim 0.132\bar{z}\chi(\bar{z}). \tag{3.28}$$

Since the linear limit of l is redshift dependent, the high l matrices will have only bins at higher redshift, and exclude the bins at lower redshift. This does not affect the likelihood expression above besides the size of the $C(l)$ matrix.

3.5 Experimental Setup

The redshift distribution of galaxies is crucial for the inference of cosmological parameters, and modified gravity is no exception. In order to self consistently constrain the photometric redshifts, and to propagate uncertainties in the redshift distribution through to the cosmological parameters, we need to sample over possible redshift distributions and marginalise the result to reduce it to just the cosmology. In this case, we shall again be sampling over the mean and standard deviation of a set of eight Gaussian redshift bins to represent the photometric sample and its uncertainties. As shown in the previous chapter, such uncertainties can be strongly constrained by cross correlation with a spectroscopic sample. In this chapter, we shall make use of a much larger and denser spectroscopic survey which more closely resembles the wealth of data one might expect to make use of in the Euclid analysis. The Euclid like parameters for this survey are based on [142], which details how much data we can expect and over what redshift ranges.

The setup for our 'experiment' involves eight photometric bins and twenty spectroscopic bins. Galaxy number counts are calculated for both photometric and

spectroscopic bins, whereas shear spectra are only calculated for photometric selections. In order to reduce the exhaustive computations required for the sampling, the l-range is capped at $l = 250$, even for bins where the linear scales may extend above this. Although the weak lensing may be calculated up to much higher $l \sim 5000$, we are predominantly interested in modified gravity, which typically acts at larger scales, and the cross correlations between number counts and lensing as a probe of deviations from GR. Since galaxy counts are restricted to l of a few hundred, and therefore so are the cross-correlations, we neglect to analyse the high l part of the weak lensing spectra. (This also makes the computational cost considerably lighter.)

The total redshift distribution is assumed to be approximately of the form [46]

$$n(z) \approx z^2 \exp\left[-\left(\frac{\sqrt{2}}{0.8}z\right)^{\frac{3}{2}}\right], \tag{3.29}$$

although the exact distribution will not be strict to this form, as it depends on the choice of bins we use as our fiducial cosmology. The relative height of each bin is chosen so that their sum roughly matches the shape of the model in equation 3.29. These relative heights in then translated into a number of galaxies per bin, so that the total number of galaxies is 1.5×10^9 for the photometric sample and 5×10^7 for the spectroscopic sample, in accordance with Euclid estimates. The number of galaxies in each bin is only needed for calculating the noise terms; where $n(z)$ enters into the window functions for the $C(l)$ calculation it is always a normalised probability distribution.

For smoothness and ease of computation we use gaussian bins everywhere, although the exact shape of the bins is unlikely to have a significant impact on the results.

3.5.1 Photometric redshift distribution

Photometric redshifts are in the range $z \lesssim 2$, and we shall split the photometry in to eight Gaussian photometric bins of uniform width and spacing in redshift. The properties of these redshift bins (the mean, variance, and number of galaxies con-

Table 3.1: Properties of fiducial Gaussian photometric redshift bins. $N_{\rm gal}$ is selected so that the heights roughly follow the shape of the equation 3.29, and so that the sum is close to 1.5×10^9, the total expected Euclid galaxy sample.

	Bin 0	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7
μ	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7
σ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$N_{\rm gal}/10^8$	1.58	2.82	3.20	2.81	2.08	1.34	0.73	0.41

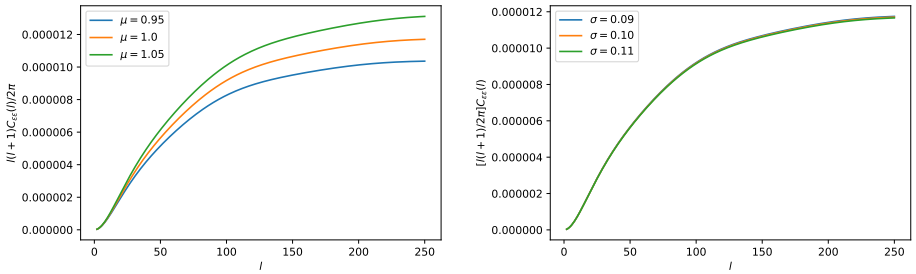
tained in each) are summarised in table 3.1. Galaxy number counts (with redshift space distortions) and weak lensing shear $C(l)$s are calculated for all of the photometric bins and their cross-correlations.

3.5.1.1 Weak lensing and $n(z)$

As we did in section 2.2.3, it is useful to briefly inspect the effects of μ and σ of the photometric bins on the weak lensing $C(l)$s. Fig 3.3 shows the impact of modifying the first and second moments of the redshift bin on the weak lensing auto-correlation. The power increases with increasing μ, since there is more material (and more structure) in between the source and the observer to lens the light. The $C(l)$ is comparably insensitive however to the width of the bin, σ. Because the weak lensing signal depends on all the matter between the source and the observer, the adjustment to the width of the bin is absolutely tiny when compared to this distance and much of the source matter remains at a similar redshift, and thus has little impact on the amount of lensing accumulated between source and observer. As a result, weak lensing may constrain μ but not σ, and better constraints for $n(z)$ are likely to come from the galaxy number counts.

3.5.2 Spectroscopic redshift distribution

Unlike in the previous chapter, we simulate a dense spectroscopic redshift sample across a wide redshift range. As before, spectroscopic redshifts are fixed and thus do not contribute to the dimensionality of the parameter space, but do have a computational time cost due to the large number of cross spectra. In this work we calculate all of the cross and auto-correlations, although cross-correlations between widely separated bins could likely be discarded without impact to the results. (N.B.



(a) **(b)**

Figure 3.3: Left: Variation of an autocorrelation $C_{\gamma\gamma}(l)$ with the mean of the redshift bin μ (σ fixed at 0.1); Right: Variation of an autocorrelation $C_{\gamma\gamma}(l)$ with the width of the redshift bin σ (μ at 1.0).

This is not true if there are unknown anomalies in the photometric redshifts such as an outlying overdensity in redshift space, see section 2.7.3.) Spectroscopic redshifts will be available from Euclid in the range $0.7 \leq z \leq 2$ for some fifty million galaxies. It is important that the width of the bins in the spectroscopic sample be less than the BAO scale ~ 100 Mpc. For this to be the case, a width of $\Delta z = 0.03$ is appropriate across the redshift range (although it can be as broad as 0.05 at high redshift). If necessary, additional spectroscopic information can be added in the low z range from surveys such as BOSS.

We have 20 redshift bins, which are arranged to roughly replicate the redshift distribution in equation 3.29 over the range $0.7 \lesssim z \lesssim 2$. These distributions are shown in Fig. 3.4, along with their sum which roughly matches the model for the total $n(z)$ distribution. The width of the bins is also increasing with redshift (with widths $0.02, 0.021, ..., 0.039$): they need to be kept small enough to be less than the BAO scale, but otherwise as wide as possible to reduce the number of bins and thus computational labour. Only galaxy number counts (with RSD) are calculated for spectroscopic bins.

3.5.3 CMB

As in the previous chapter, we use CMB-TT information up to $l = 3000$ and assume a Planck-like noise function in order to constrain parameters which are otherwise poorly constrained by galaxy data. Cross correlations between the CMB and the

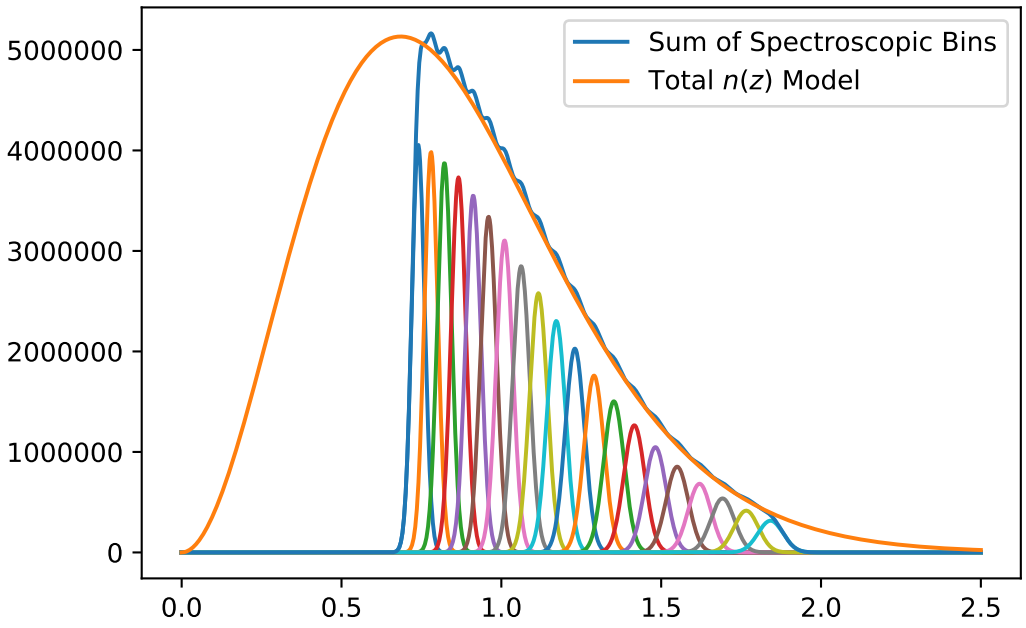


Figure 3.4: Redshift distribution of galaxies for the spectroscopic sample. This sum of the bins is compared against the redshift model in equation 3.29.

galaxy samples are not currently calculated. The details of the CMB information can be found in section 2.4.2.2.

3.5.4 Noise functions

The noise for galaxy number counts and for the CMB is modelled in the same way as in section 2.4.2, but in this case the fraction of the sky covered is much greater, and we take $f_{\text{sky}} = 0.36$ (from $A \approx 15,000\,\text{deg}^2$). The number of galaxies per photometric bin can be found in table 3.1.

For the weak lensing, we use the noise function [46]:

$$N_\gamma = \frac{2\pi f_{\text{sky}} \sigma^2_{\text{shape}}}{N_{\text{gal}}} \tag{3.30}$$

and we take $\sigma_{\text{shape}} = 0.35$ in accordance with Euclid estimates. Similarly to galaxy number counts, this noise is only added on the diagonal (i.e. autocorrelations).

3.5.5 Fiducial Model and Cosmological Parameters for Sampling

We sample over a limited set of cosmological parameters, $\{A_s, \Omega_b, \Omega_{cdm}, h_0, w, n_s\}$, using flat backgrounds only; the fiducial model for the likelihood is taken to be ΛCDM.

We run two sampling chains in this chapter. The first, predating the recent GW observation, includes α_T, and uses a ΛCDM background. The modified gravity parameters sampled over are therefore $\{\alpha_K, \alpha_B, \alpha_M, \alpha_T, M_*^2\}$. The second is extended to include wCDM to model the possible changes to the expansion history through dark energy but, as it was launched after the GW observation, also excludes α_T. The modified gravity parameters are $\{\alpha_K, \alpha_B, \alpha_M, M_*^2\}$.

The ΛCDM limit is when all $\alpha_i \to 0$ together, and $M_p = 1$. The numerical calculation of the $C(l)$s in this limit was checked against the ΛCDM $C(l)$s to ensure that the likelihood was behaving well.

3.6 Sampling Results

In the results presented here we will typically marginalise over either the cosmology or the redshift parameters in the plots for clarity, since there are so many parameters in the analysis. As well as forecasting both MG constraints and precision $n(z)$ for Euclid, we are in a position to investigate what has been changed phenomenologically by removing α_T from our consideration. (Or, put more precisely, fixing $\alpha_T = 0$.)

We find that the modifications to gravity in a ΛCDM background show some degeneracy with cosmological parameters via α_B and α_M. In the run including a wCDM background the modified gravity parameters exhibit their strongest degeneracy with w; this in turn leads them to have similar degeneracies with the other cosmological parameters as w does, in some cases reversing the trend seen in the ΛCDM analysis (see red and blue contours in fig 3.6 for wCDM and ΛCDM respectively).

Unsurprisingly, the initial value of the Planck mass is highly degenerate with

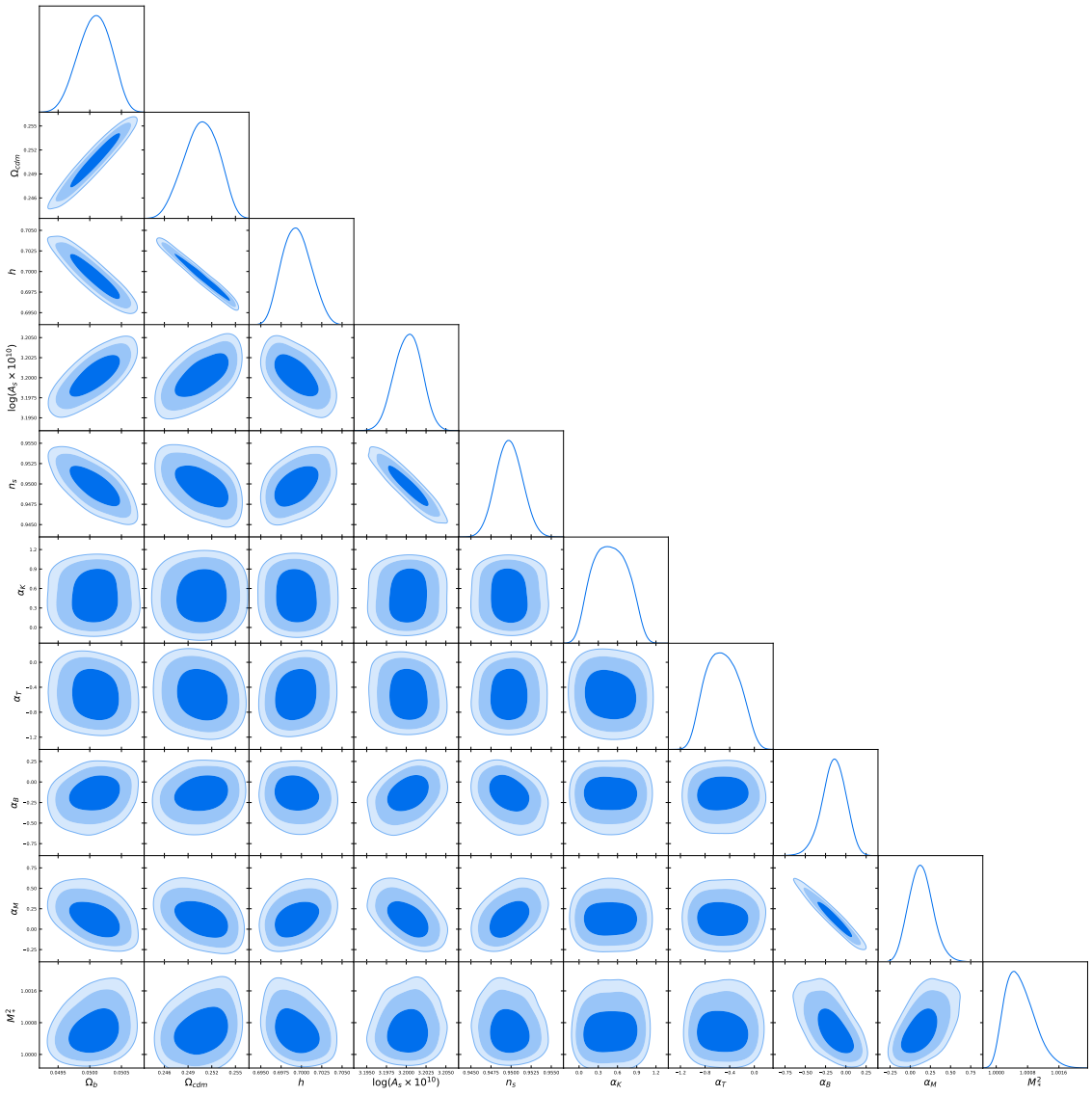


Figure 3.5: Triangle plot showing the posterior contours for the standard cosmological parameters and the five modified gravity parameters where α_T is free to vary. Neither α_K nor α_T are constrained within our prior range ([0,1] and [-1,1] respectively), except that a strong constraint on $\alpha_T \leq 0$ from stability. α_M and α_B show some degeneracy with the cosmological parameters, especially A_s and n_s suggesting that they are combining to affect the amplitude and tilt of the power spectrum. Most strikingly, we have a tight probability contour around the line $\alpha_M = -\alpha_B$, which permits models with $f(R)$ like properties for these α-functions.

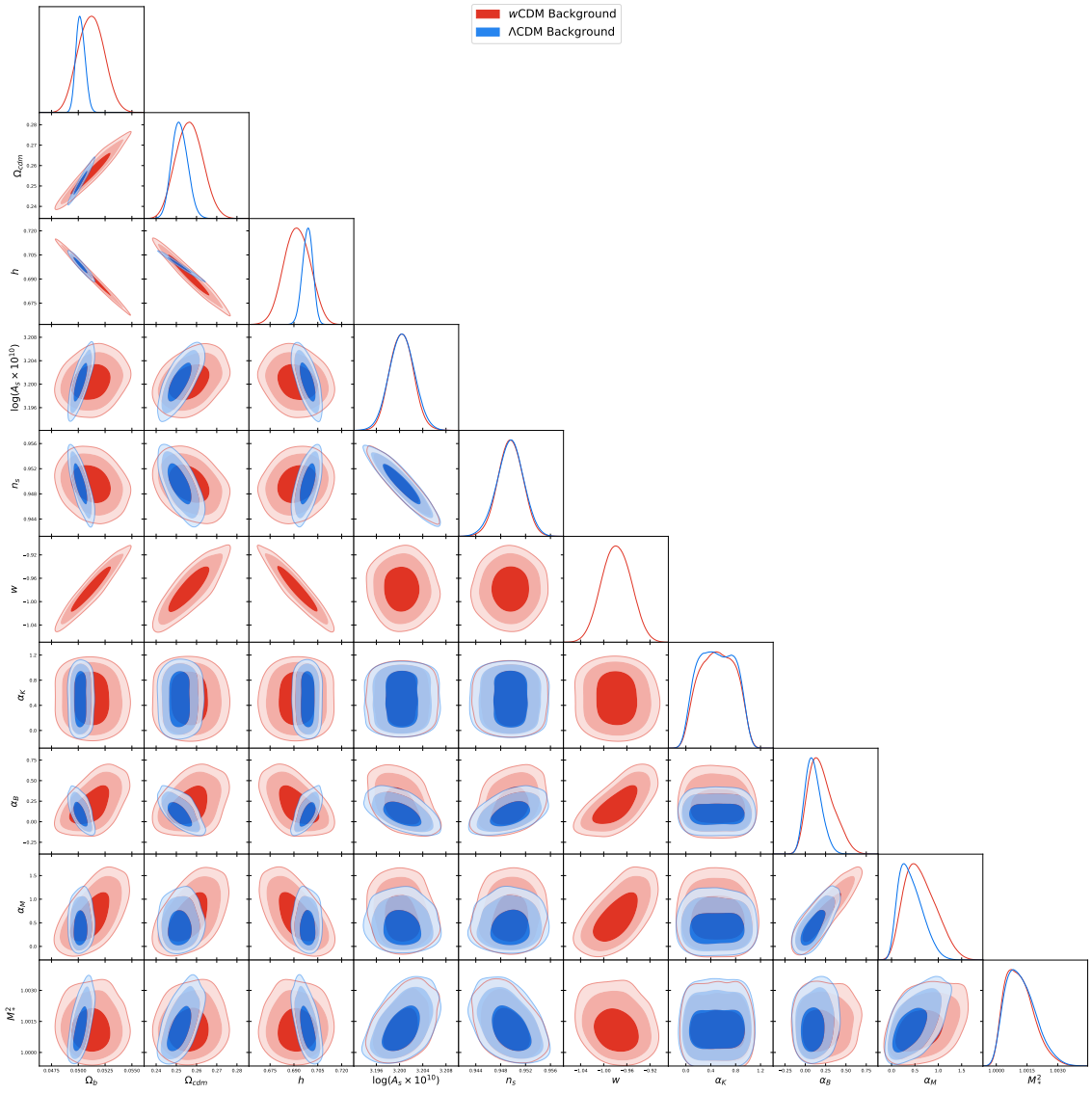


Figure 3.6: Triangle plot showing the posterior contours for the standard cosmological parameters and the four remaining modified gravity parameters (after α_T has been excised). We note that the cosmological constraints are much weaker due to the introduction of w which is highly degenerate with many of the parameters in our analysis. α_M has less visible degeneracies with standard cosmological parameters, although there is some degeneracy with w in particular. α_B now shows marked degeneracy with A_s and n_s. Comparatively poor constraints on M_*^2 may be causing the α_M contour to extend much further, or vice versa, as they are highly degenerate.

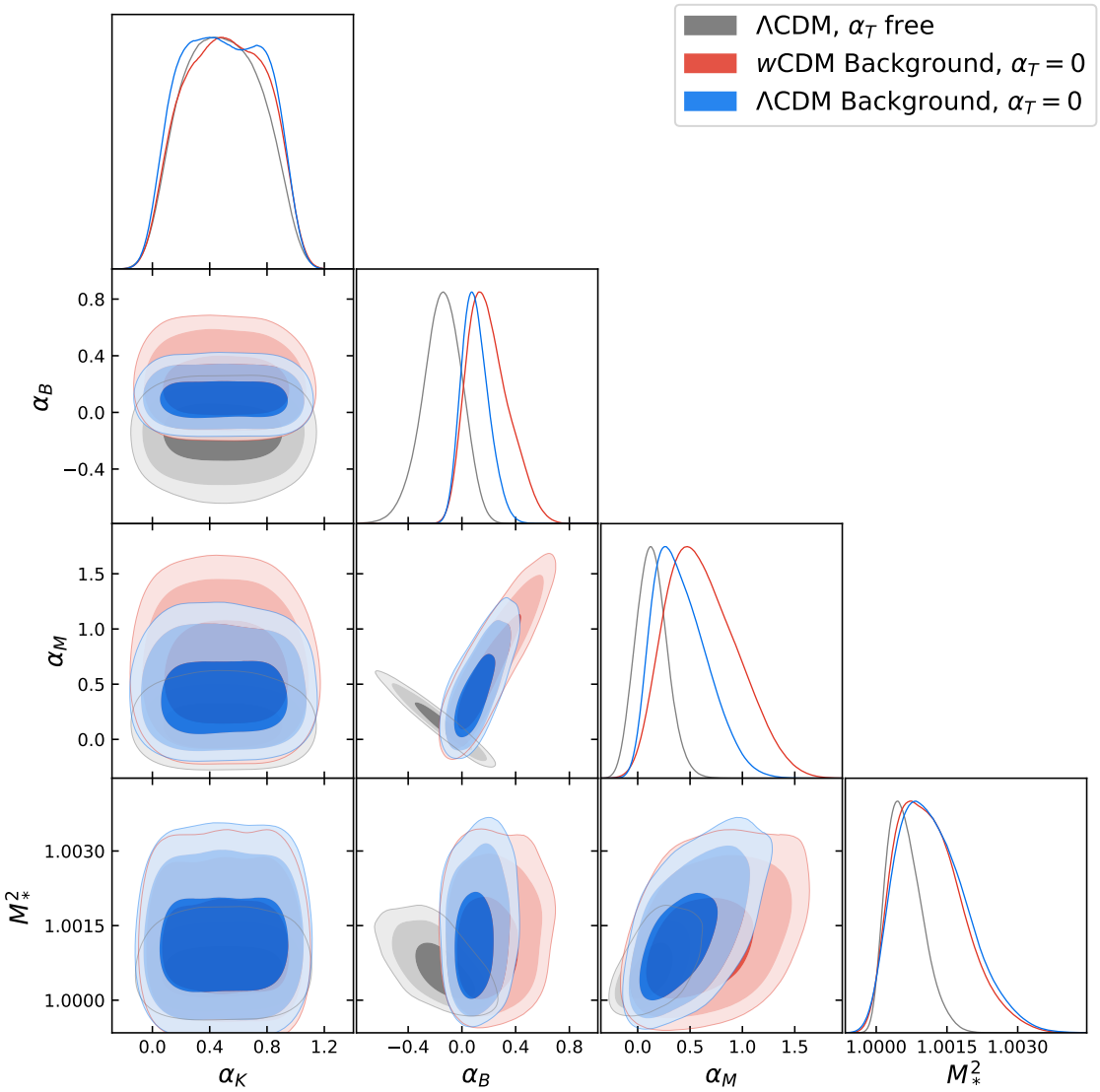


Figure 3.7: Triangle plot showing the posterior contours for the four modified gravity parameters shared by the analyses. α_K is not constrained by the analysis, but is allowed to vary in the range $[0, 1]$ because it can interact slightly with other parameters, such as the setting of the 'braiding scale'. The constraints on α_M and α_B are very different depending on the inclusion of exclusion of α_T as a free parameter, particularly with respect to their degeneracy. Closing off the area $\alpha_M \approx -\alpha_B$ forces the α-functions to be positive due to a highly asymmetric likelihood and stability conditions on either side of this line. In the case where $\alpha_T = 0$ the degeneracy with M_*^2 is particularly strong, and the constraints on α_M are in fact weakened.

its running parameter α_M, although they are positively correlated (a large initial mass with an positive mass running); we have already seen in section 3.2.3.1 that these two can combine to produce μ close to 1. The kineticity α_K, which only affects things on very large scales, remains unconstrained within in the limits of the prior in this analysis. It is possible that it could be constrained from taking very large values $\alpha_K \gtrsim 10$. Nevertheless we allow α_K to vary in the range $[0,1]$. This is because α_K can subtly influence aspects of the model such as the scale dependence of α_B-effects, and in general with these models it is a good idea to avoid getting too stuck in one part of the parameter space in case it has pathological features. As we shall see, setting $\alpha_T = 0$ changed the viable parameter space enormously, despite itself being very poorly constrained!

3.6.1 Redshift constraints and degeneracies

The mean redshifts and variances of the bins are extremely well constrained across the range, despite the fact that the spectroscopic sample begins at higher redshift than the photometry. Although the lowest redshift bin, which has effectively no overlap with any of the spectroscopic bins, has a poorer constraint than the others, it is nevertheless reasonably well constrained by its overlap with the neighbouring redshift bin, having the same uncertainty on μ as the redshift bins in table 2.3 in the previous chapter. If stronger constraints are required at low redshift, the BOSS catalogue can be combined with the Euclid spectroscopic set to fill in the low end. Compared to our previous analysis (see table 2.3) we have slightly improved constraints on bins towards the middle of the range by a factor of 2-3 on μ, and up to a factor of 2 on σ. Clearly there are diminishing returns on increasing the size of our spectroscopic sample; nevertheless, the dense spectroscopic dataset is valuable in its own right as a measure of cosmology, and to catch any unusual shapes or artefacts in the redshift distributions which might be missed by sparser sampling (see section 2.7.3).

The degeneracies between the redshift parameters and standard cosmology was discussed in the previous chapter, and once again we have constrained redshift to below a level where it is visibly degenerate with cosmology. The modified gravity

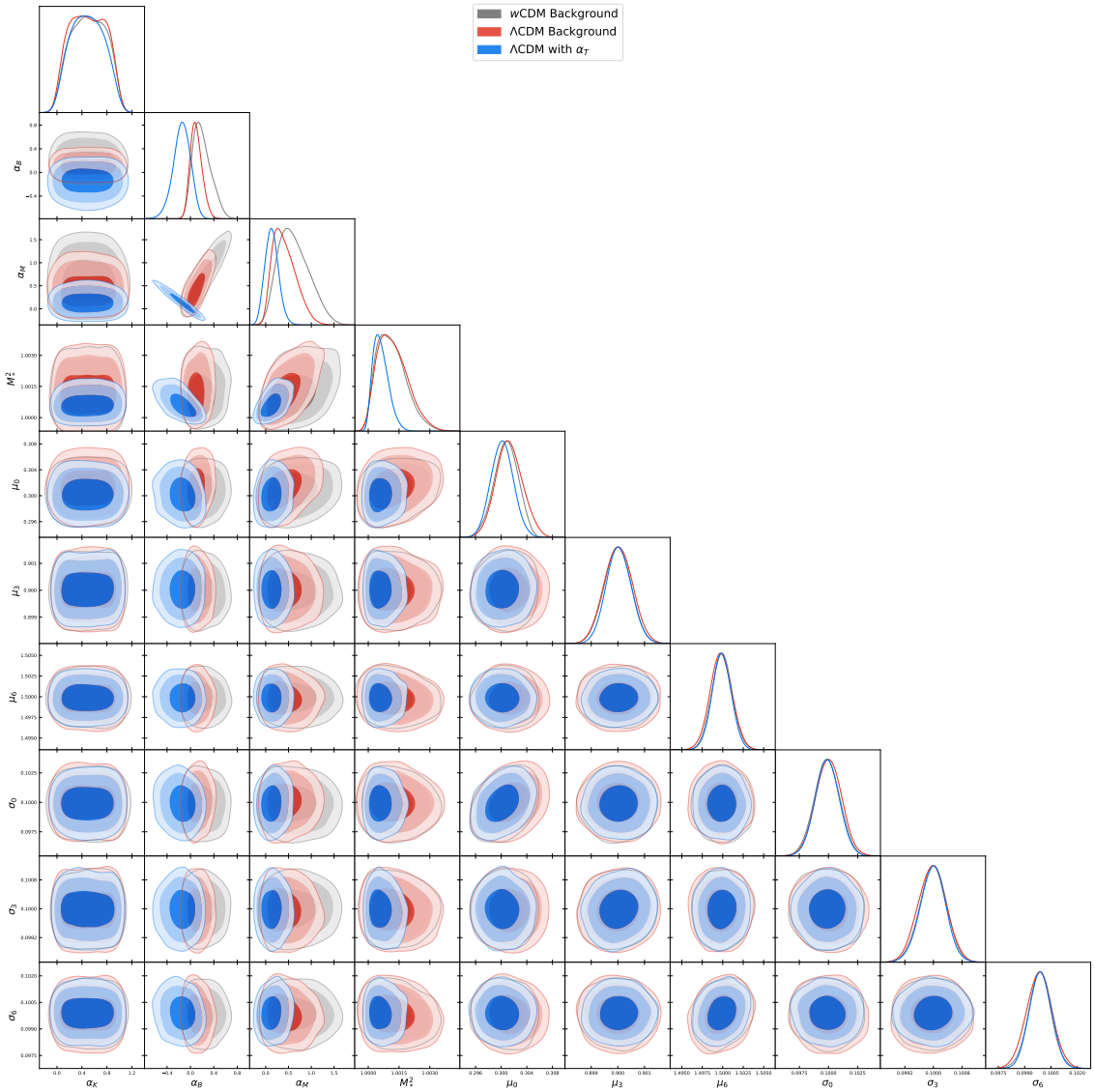


Figure 3.8: Triangle plot showing the posterior contours modified gravity parameters (excluding α_T) for all three runs alongside a representative subset of the photometric redshift parameters. The redshift parameters are largely independent of the α-functions as well as each other at this level of constraint (although some mild degeneracies still exist). Photometric redshift constraints are very similar between all runs, with a slight increase in uncertainty when w is introduced.

Table 3.2: Constraints on $n(z)$ at 68% confidence, using cross-correlations between galaxy number counts and shear in photometric bins with galaxy number counts in spectroscopic bins. Results are taken from the run with $\alpha_T = 0$ and including w; constraints on $n(z)$ are very similar between the two runs.

	Bin 0	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7
$\Delta\mu \times 10^3$	± 3	± 2	± 1	± 0.6	± 1	± 1	± 1	± 2
$\Delta\sigma \times 10^3$	± 1.0	± 0.9	± 0.5	± 0.5	± 0.7	± 0.7	± 0.9	± 1.0

parameters show little degeneracy with any of the photometric redshift parameters in any of the runs, as shown in fig 3.8. (The mean and width of the lowest redshift bin has a mild degeneracy with α_M. This might well vanish is the low redshift range were constrained as well as the mid to high range.) This does not mean that redshift errors would not impact MG analyses, rather that the redshifts have been constrained below the level of significantly influencing the results by the information contained within the cross-correlations, as in the previous chapter with ΛCDM. The fact that we can achieve this within a modified gravity analysis is very promising, as biased $C(l)$s might lead to spurious detections (or non-detections, as the case may be).

3.6.2 Modified Gravity and cosmology with and without $\alpha_T \neq 0$

In this section and the next we shall marginalise over all redshift parameters, and only consider cosmology. This first run allows all of the α-functions to vary freely. The constraints, shown in fig 3.5 and enlarged in fig 3.7 in grey, on α_B and α_M are competitive with other analyses [138][137], although both α_K and α_T are largely unconstrained by the observables that we have chosen. In the case of α_K this is to be expected, since it only affect the very low k part of $P(k)$, and only noticeably so for very high values of α_K not considered in this analysis.

One of the most striking features is that the model in the $\alpha_B \times \alpha_M$ plane is tightly constrained to lie along $\alpha_B \approx -\alpha_M$. The archetypal model of this kind is $f(R)$ gravity, although $f(R)$ has a different time dependence, and sets $\alpha_K = \alpha_T = 0$. This suggests that such a theory as $f(R)$ would be difficult to differentiate from ΛCDM with our given experimental setup. Looking at papers such as [138] and [137], and even at the two analyses in this paper, it is clear that different setups – both experimental and theoretical – can yield very different degeneracies. This is promising, as it suggests that combination of Euclid and CMB-TT with other datasets could be much more constraining than they are in this forecast.

The analysis with $\alpha_T = 0$ on ΛCDM and wCDM backgrounds, fig 3.6 and in fig 3.7 in blue and red respectively, is very different indeed. The constraints on the cosmological parameters are wider in the wCDM case, predominantly due to

degeneracies with w introducing uncertainty. Even more dramatic is the change to the MG parameters. $\alpha_B \approx -\alpha_M$ is now strongly disfavoured, despite being a strong constraint in the previous run. Looking at the results from the previous run, α_T is not obviously degenerate with these parameters (or anything else), so one might wonder how it is having such an impact. In fact, α_T is vital for stabilising large areas of the parameter space when the fiducial time dependence for the α-functions is employed, although the parameter itself has little impact on the observables. The degeneracy between α_B and α_M now runs in almost the orthogonal direction to before. Understanding the stability of these models will be key to understanding this strange behaviour.

3.7 The role of α_T and parameterisation

Despite the fact models with non-zero α_T are generally considered to be ruled out by gravitational wave measurements, we find in this analysis that α_T plays a crucial role in our ability to explore the theory space. When α_T is permitted to be non-zero, it has an important stabilising effect on c_s^2, even though the tensor excess itself does not have much direct impact on the observables (and thus is very weakly constrained). In this case, we find that $\alpha_M \approx -\alpha_T$ is strongly favoured. What is interesting is that, when $\alpha_T = 0$, this area of the parameter space is excluded on the basis of stability criteria. So although α_T is poorly constrained in its own right, removing it drastically changed our constraints on the other Horndeski parameters.

This change however is highly concerning. The instability of $\alpha_B = -\alpha_M$ is not generic, but a characteristic of their time dependence. $f(R)$ gravity, for example, is a family of theories which can produce stable cosmologies, and satisfied $\alpha_B = -\alpha_M$ and $\alpha_T = 0$. In the parameterisation chosen in this paper – which is quite standard in the literature – the balance of α-functions of $f(R)$ gravity is unfairly ruled out, based not on the effect on the observables but on model pathology.

Such problems call into question the usefulness of parameterising the Horndeski class in such a way. One of the appeals of the method is the connection to the underlying action, and the physical properties of the theory. The process of recover-

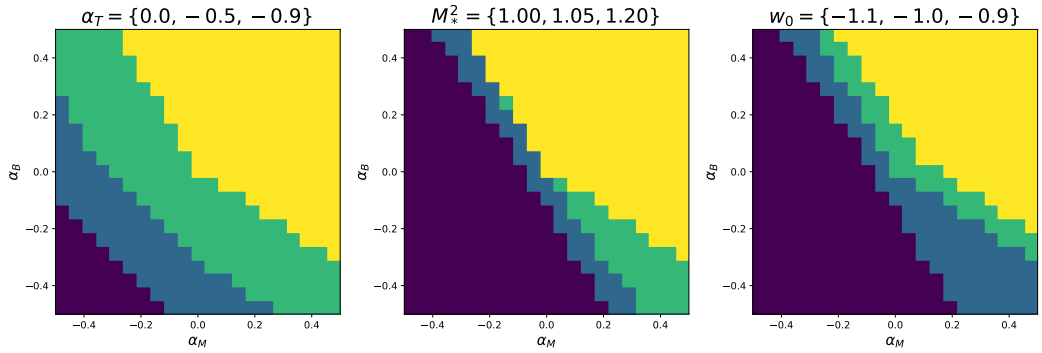


Figure 3.9: Maps of the exclusion zones in the $\alpha_M \times \alpha_B$ plane due to gradient instabilities. In the leftmost plot the yellow region represents $\alpha_T = 0$, and subsequent bands are decreasing α_T. The darkest band is unstable for all three values. The background is ΛCDM and $M_*^2 = 1.0$. In the middle plot, the background is again ΛCDM but $\alpha_T = 0$. The yellow area is stable when $M_*^2 = 1.0$. Subsequent darker bands show areas which are stable when $M_*^2 = 1.05$ and 1.20 respectively. The darkest band is unstable for all of these values. The rightmost plot fixes $\alpha_T = 0$ and $M_*^2 = 1$, but allows for a wCDM expansion. The yellow region represents the stable region for $w = -1.1$, which subsequent regions corresponding to increasing w. Again, the darkest region is unstable for all values of w considered.

ing the α parameters based on a simple and strict time dependence strongly biases the constraints one obtains. Functions which are allowed to have different and independent shapes from one another may be much more stable. An interesting question to ask is, if given a mock data set generated on $f(R)$ cosmology, what cosmology would this fiducially parameterised analysis return? The implied structure of the action would necessarily be incorrect since that region is excluded by stability, but what balance of parameters might the analysis find to compensate? And would it find a modified gravity theory at all or, given that $f(R)$ models are not available to it, would it find that there is no improved fit from GR, thus missing the evidence for deviations entirely?

3.7.1 Stability of the models in $\alpha_i \propto \Omega_{DE}$

The HICLASS code checks the models for some basic stability criteria; we have already discussed to some extent the importance of the gradient instabilities where

$c_s^2 < 0$. The full criterion (including α_T) is [153]:

$$c_s^2 = -\frac{(2-\alpha_B)\left[\dot{H} - H^2(\alpha_M - \alpha_T) - \frac{1}{2}H^2\alpha_B(1+\alpha_T)\right] - H\dot{\alpha}_B + \frac{8\pi G\rho_m}{M_*^2}}{H^2\left(\alpha_K + \frac{3}{2}\alpha_B^2\right)} \quad (3.31)$$

This stability criteria is the one most often violated when sampling the hypercube of α-parameters. Fig 3.9 shows maps of the stable regions for different model parameters. We can see immediately that α_T is highly effective at removing gradient instabilities, which can be easily understood by inspecting the dark energy sound speed equation. Although it opens up the parameter space considerably, α_T has little other effect on the observables in this setup, and thus is therefore poorly constrained. In this way, it allows this area of parameter space to be explored without greatly reduced likelihood. Indeed, the best fit in the $\alpha_M \times \alpha_B$ plane lies along the line $\alpha_B = -\alpha_M$, which is intrinsically unstable for $\alpha_T = 0$, $M_*^2 = 1.0$ (except at the point $\alpha_B = \alpha_M = 0$). Once α_T is removed from the analysis, this parameter space becomes much less accessible. It can be partially recovered by moving M_*^2 and w_0 away from their fiducial ΛCDM values, but this comes with a heavy penalty to the likelihood, since they are stringently constrained. As a result we find that the shape of the modified gravity constraints are drastically changed by the removal of α_T.

Looking at fig 3.9 also goes some way to explaining why our $\alpha_M \times \alpha_B$ contours move to the place that they do. Although in the $\alpha_T \neq 0$ analysis the contour appears roughly symmetric around the line $\alpha_M = -\alpha_B$, we can see that further away from the line, and without α_T to provide stability, this region is highly asymmetric in its likelihood properties.

3.8 Discussion

The constraints obtained in this analysis suggest that Euclid's large galaxy catalogue should be a promising tool in constraining deviations from GR. The poor performance of the observables in constraining α_T might have been cause for concern, but this has been neatly taken care of by gravitational waves. (One might be

inclined to leave α_T in one's analysis if one had motivation for a *decreasing* α_T which could vanish at late times.) Combination with other probes would be vital for the constraint of this parameter. α_T appears to be tightly constrained in [137]. That study uses more cosmological probes – CMB temperature, lensing, and polarisation, BAO, RSDs, and $P(k)$ – and finds broader constraints on α_B and α_M. Interestingly this analysis does not contain the ΛCDM limit point a the 2σ level, and α_T is constrained to be at the very bottom of its allowed range. This stands in stark contrast to the new constraints on α_T from gravitational waves. Given the significant change to our dataset from removing α_T, such existing constraints may no longer be a reliable indication of the state of observation Horndeski gravity. Nevertheless, this dataset clearly had access to constraining power that we do not have, and thus in a future Euclid analysis it would almost certainly be beneficial include the CMB lensing and polarisation data as well. Forecasts for LSST combined with other probes including a hypothetical stage-IV CMB experiment in [138] are comparable to ours for α_B, but much stronger for α_M. Of the datasets included in that analysis, the strongest source of constraint on α_M comes from the LSST catalogue. These results are not directly comparable with our own however, since they use a simplified Fisher matrix analysis (which assumes that the likelihood is perfectly ellipsoidal and thus determined by the gradient at the fiducial point), and the fiducial cosmology is not ΛCDM but includes non-zero α-parameters. Again, the exclusion of α_T may produce interesting effects on these forecasts, and we have seen that in preparing thorough forecasts for the next generation(s) of experiments, we will need to be very careful about how we parameterise our models, and what their time dependence is.

We have seen that a dense spectroscopic sample can be finely binned and included in a sampling analysis with photometric bins for a realistic survey setup. The cross-correlations between photometric and spectroscopic bins yield constraints of $\mathcal{O}(10^{-3})$, with the best constraints being in the centre of the sample. Those bins at the fringes which have no overlap with spectroscopic bins are still constrained to the same order of magnitude by the correlations with their neighbouring bins,

which are well constrained. It is important to cover as much of the redshift range as possible, as the further we get from the anchoring of the spectroscopy, the more the uncertainties can accumulate down the line. In this case, it may be better to arrange spectroscopic samples covering the entire redshift range but with small gaps than to cover a partial area very closely, if one is worried about the computational cost of including too many bins. The fact that the redshift bins do not show any strong degeneracies with parameters in the cosmological or MG parameter sets suggests that this method is sufficiently constraining to be put to use in the next generation of surveys.

In order to proceed with next generation surveys, we will need to develop more robust, or better motivated, models of dark energy which we can sample with confidence. The question of how effective parameterisations are at recovering full modified gravity models remains open, and should be carefully considered in future approaches to observational modified gravity and dark energy in order to reliably understand the constraints that we are obtaining in our forecasts or, in the future, in our analysis of data. This forecast may be extended both by exploring more flexible models, and by broadening our dataset to include the neglected CMB polarisation, CMB lensing, weak lensing shear at high l, and supernovae.

Part II

The Local Group

Chapter 4

The Mass of the Local Group

Of course, Andromeda did
everything the Milky Way did,
backwards in high heels!

---

not Bob Thaves

4.1 Introduction

In this section we will briefly turn our attention away from physics on the largest scales, and investigate the mass of our own home, the Local Group of galaxies. We shall see that this question is not so widely separated from cosmology as one might think! Precise estimates of the mass of the Local Group (LG) remain an outstanding problem for both observers and theorists, with physical properties being difficult to observe and reliable models remaining elusive. Recent interest has been fuelled by the advent of large N-body simulations and the introduction of dark energy models. We introduce here a new approach to the problem by using machine learning to model complex phenomena present in the simulation.

For over fifty years, the timing argument (TA), introduced in [92], has been used as a simple dynamical estimate for the the mass of the LG, assuming that the mass is dominated by the Milky Way and Andromeda. The argument uses simple Newtonian mechanics to calculate the combined mass of the halo pair.

The various mass estimates of the LG, some of which are discussed and compared in [122], demonstrate the lack of consensus in determining the LG mass. In

order to build a coherent picture of the near universe, and how it fits into the universe as a whole, we require robust estimates of the LG mass and the ability to rule out particular models. Far from being an isolated system, the LG sits within a filament of the vast cosmic web, and its history is cosmology dependent through the expansion of the universe and the formation of structure. Given the dependence on Λ and large scale structure, we must work towards a means of utilising a variety of aspects of cosmology. A better understanding of the mass of the LG and other systems would help us to evaluate which models are successful in galaxy formation, dynamics, and near field cosmology.

A number of more detailed models have been made to estimate the mass of the LG. [112] includes observations on smaller bodies (other than MW and M31) in the LG. The introduction of a Cosmological Constant has also been explored ([80], [117], and references therein), which manifests as an additional expansion term and increases the mass estimate for the LG by 13%. The use of data from large cosmological simulations was introduced by [95] and furthered by [98], which allowed for an analysis of the robustness of the TA, and calibration of the TA mass estimate to a simulation mass to resolve biases. Li & White find that the TA is an (almost) unbiased estimator of the mass of an LG like system, based on a selection of halo pairs selected for their similarity to the LG from the Millennium simulation, and assuming no effect from a cosmological constant. A more recent study by [90] uses simulations to calculate a likelihood for the masses, and [81] uses a constrained simulation approach to generate LG like objects and study their mass distribution.

From the previous work done estimating the LG mass, it seems clear that there is still a large uncertainty in predicting the simulation mass ($M_{\rm sim}$) using available analytic models. Since these models approximate the galaxy as a point mass, and we know that galaxy and the dark matter halo which it occupies is an extended mass distribution, the mass estimate $M_{\rm TA}$ is not necessarily inherently physically meaningful. From a practical point of view we would like to have an estimate of more physical mass, such as the virial mass or M_{200} (the mass contained within a radius r_{200} such that the density within r_{200} is a factor of 200 greater than the mean

density of the universe).

In this work we combine data from N-body simulations with machine learning methods to exploit information about the environment of the LG within larger scale structure to improve estimates of the LG mass in the context of ΛCDM cosmology. We use Artificial Neural Networks (ANN), a class of machine learning algorithms which have been widely applied in other areas of physics. The network is trained by providing it with the salient physical parameters and masses for halo pairs selected from the simulations; the algorithm seeks a function which best predicts the output (combined halo mass) from the inputs (observable dynamical and local environmental parameters). Once it has converged on a solution, the relevant inputs for the LG (from observations) are fed into the function, which then returns a mass.

In the next section we describe the simulation used, and the criteria for identifying and selecting halo pairs. In section three we review some of the TA models which have been used in the literature, apply them to our simulation data set, and look at the robustness of the assumptions of the TA. In the fourth section we give a brief introduction to artificial neural networks, and apply these techniques to mass estimation in the simplest case of only utilising separation and radial velocity information. In section five we explore additional parameters and cuts to the dataset to refine ANN estimates. In section six we apply our models and ANN to the case of the LG itself.

4.2 Simulations & Selection Criteria

The selection of Halo pairs and extraction of the environmental information such as the smoothed shear, tidal, and density fields was done by Dr. Noam Libeskind. The simulation data is taken from the *Small MultiDark Planck* (SMDPL) simulation, downloaded from the publicly available cosmosim database (www.cosmosim.org, [123]). The simulation box size is 400Mpc/h (h = 0.6777), with 3840^3 particles, with a particle mass of $9.63 \times 10^7 M_\odot/h$ and a force resolution of $1.5\text{kpc}/h$ ([94]). Halos are identified using a friends-of-friends algorithm ([93]). Suitable halo pairs are selected as follows:

1. Candidate halos are selected with a mass $5 \times 10^{11} M_{\odot} \leq M \leq 10^{13} M_{\odot}$.

2. If the candidate is between 1.5 Mpc and 3 Mpc of a another halo of mass $> 10^{12} M_{\odot}$ then the candidate is discarded.

3. If the candidate is within 0.5 Mpc of a another of mass $> 5 \times 10^{11} M_{\odot}$ then the candidate is discarded.

4. If there is another candidate (i.e. halo with mass $5 \times 10^{11} M_{\odot} \leq M \leq 10^{13} M_{\odot}$) at a distance 0.5Mpc$\leq r \leq$1.5Mpc then the pair is accepted.

Our criteria are less restrictive than those used in many other studies, such as [98], because we also wish to investigate broader applicability of the TA outside of the LG. We have in total 30190 halo pairs. This data includes the position, velocity and mass of each halo, as well as environmental data such as the local density, shear (see equation 4.4), and velocity fields. We can easily calculate dynamical parameters (r, v_r) from $r = |\mathbf{x_B} - \mathbf{x_A}|$ and $v_r = (\mathbf{v_B} - \mathbf{v_A}) \cdot \hat{\mathbf{r}}$, as well as the tangential velocity v_t.

4.3 The Timing Argument and its Extensions

In this section we will discuss some of the variants of the TA that have been proposed, and compare their performance on our dataset. These provide an important benchmark for estimators of the binary mass, as these still form a basis of our understanding of the LG mass. This is the first time different TA-like models have been systematically compared on a simulation data set, allowing us to select the best TA-based estimator. The results can be seen in Figure 4.1 and section 4.3.1.

TA predicts mass as a function $M_{TA} = M(r, v_r, t_u)$, where r is the separation between galaxies (or halos), v_r is radial velocity, and t_u is the age of the universe (i.e. time at current epoch assuming $t = 0$ at "big bang"). Given t_u from experiments such as Planck and the relevant cosmological model (we shall use $t_u = 13.8$ Gyr [121]), we may write $M_{TA} = M(r, v_r)$. The assumptions made by the TA are:

1. Halos have no transverse (non-radial) velocity.

2. Halos can be modelled as point masses.

3. Halo pairs are isolated; there is no external gravitational field.

4. Halos start their orbits in the early universe close to $(r,t) = (0,0)$.

Under these simple conditions the system evolves as

$$\frac{d^2r}{dt^2} = -\frac{GM}{r^2} + \frac{\Lambda c^2 r}{3} \tag{4.1}$$

where in most treatments (and the original formulation) one takes $\Lambda = 0$. (See section 1.1.2.3 for details.) The simple form with no Λ term has a well known cycloid solution which may be found in [92] amongst others.

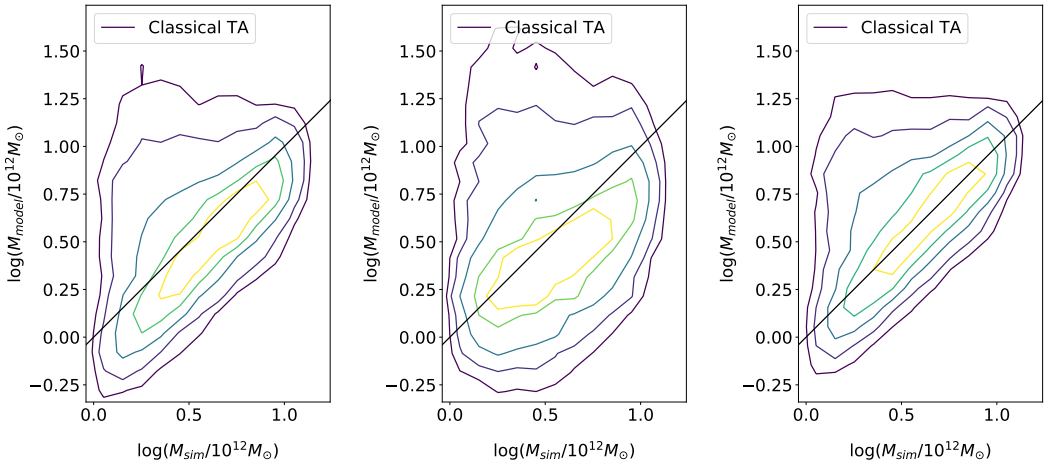
This can be extended to include transverse velocities by treating the separation and velocity as vectors, and there is an analytic solution for this case presented in [87]. Intuitively including tangential velocity might increase the mass estimate, but because the boundary condition is different (periapsis at $t = 0$ as opposed to $r = 0$ at $t = 0$) it can in fact raise or lower the mass estimate. The results of applying this method can be compared to the traditional TA in Figure 4.1.

Another simple extension to the TA is to include a non-zero Cosmological Constant [117]; unfortunately, this equation no longer has a simple parametric solution and must be solved numerically. This theory results in an estimated mass increase of about 13% for the Local Group parameters.

4.3.1 Results of applying the timing argument

To obtain benchmark results for estimators of the halo mass, we may simply apply the TA models to the set of halo pairs. As we can see from Figure 4.1a, the TA manages to match a basic trend in the data, but is unsuccessful in a wide range of low mass halos whose masses are drastically overestimated.

The correlation between the TA estimate and the bound mass is strong in the innermost contours, although there is a visible bias when using the traditional TA (Figure 4.1a). This bias is not remedied by the inclusion of transverse velocity to the TA (Figure 4.1b). Including a Cosmological Constant appears to be effective



(a) Traditional TA **(b)** Angular Momentum **(c)** With Lambda

Figure 4.1: The results of applying the TA (left), TA with angular momentum (middle), and TA with a Cosmological Constant (right) to the data set, with the simulation mass plotted on the x-axis and the estimation plotted on the y-axis. We can see that the bias in the right panel is greatly reduced compared to the left. The middle panel shows that including v_t (without Λ) does not produce an improvement, and in fact widens the contours. This may be because the tangential component of the velocity is not primordial, but acquired at later times by interaction with larger scale structure or tidal fields. Contours show the shape of a 2D histogram, with all the pairs plotted by their simulation mass and predicted mass. The contours are drawn by ordering the grid by number density of halo pairs; grid points are then added from highest density to lowest density and contours drawn at intervals where they have reached 20%, 40%, 60%, 80%, and 90% of the pairs in the sample; over plotted is the equality line $M_{\rm sim} = M_{\rm model}$.

in removing this bias (Figure 4.1c). This is an important inclusion because the additional term from the Cosmological Constant models the expansion term which is present in the simulation. This highlights a dependency of the estimate on the cosmological model: for a ΛCDM model we must include this correction to achieve a best estimate. This may be complicated further if one wishes to consider dynamical Dark Energy models such as scalar field models or modified gravity theories.

The primary problem is that for low mass halos the TA predicts a very wide range of masses including some of the highest estimates (Figure 4.1). These account for a large fraction of the total population, with 40% of the halo pairs being outside the three innermost contours which show a strong trend. Thus far we do not have a way of identifying a priori whether or not our galaxy lies in this region, so the

uncertainty in the TA remains very large.

We may calculate the rms scatter and Pearson product-moment correlation coefficient for this data, which provide a benchmark for further estimators which we will develop. The Pearson product-moment correlation coefficient can be calculated for a sample of data using the following equation

$$p = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{[(\sum_{i=1}^{n}(x_i - \bar{x})^2)(\sum_{i=1}^{n}(y_i - \bar{y})^2)]^{\frac{1}{2}}} \tag{4.2}$$

For the TA_Λ estimate we find that $p = 0.32$ when we look at $\log(M_{TA\Lambda})$ and $\log(M_{\rm sim})$. The rms scatter is 0.41 in the log mass, which would correspond to a multiplicative factor of roughly 2.6 in the mass itself.

The large scatter between the TA estimates and the simulation masses is a concern for anyone wishing to use it to estimate dynamical masses, e.g. [126] and [117]. We are well aware of the shortcomings of the TA models, despite its many extensions; we will see if we can improve this using machine learning methods.

4.4 A Brief Introduction to Artificial Neural Networks & Their Application to Mass Estimation

In addition to the simple models outlined in section 4.3, there are many complications to the history of galaxies compared to the simple Newtonian paths considered. Halos may form at different times, accumulate or shed mass in complex ways over time, or interact significantly with external tidal or gravitational fields. Rather than constructing ever more complicated analytic models from first principles, which may lead to long or unstable calculations with ambiguous boundary conditions, we instead explore the possibilities of capturing some complex behaviour by using machine learning techniques.

We wish to find a function which will calculate the mass, M, of a pair of halos based on ΛCDM simulations, given some input parameters; these include information about the dynamics of the system such as (r, v_r, v_t), as well as parameters which characterise the environment such as local density or shear information. The

problem is therefore a (non-linear) regression problem, in which we require an algorithm to find a best fit model to our simulation pairs from our inputs. ANN are one such class of machine learning algorithms, which model unknown functions using a combination of compositions of (typically) sigmoid functions. The particular program that we use is called *ANNz*, an ANN code developed by [82] for the purposes of predicting photometric redshifts (although the technique is general). The composition of weighted sigmoid functions is unintuitive, and makes their output more difficult to interpret, but means that they are much more flexible with respect to the kinds of functions that they can model, and do not require any special knowledge about the problem which might be needed in order to choose an appropriate set of basis functions. In our case we do not have an obvious set of basis functions which could be used since we have no clear idea of how our additional parameters will affect our mass estimates in terms of functional forms. Thus we may resolve that ANN are a good starting point for modelling this behaviour.

4.4.1 An overview of ANN

An ANN consists of l layers, each of which has n_k nodes ($k \in [0, l-1]$). The function is defined by this structure and a set of weights w_{ij}^k, where k runs over the number of layers, i runs over the nodes in the current layer $[0, n_i - 1]$, and j runs over the nodes in the next layer $[0, n_j - 1]$. Each node has a value associated with it, u_i^k. Starting from the inputs, which form the first layer u_i^0, at each node a sigmoid activation function is defined as $g(u_i^k) = \frac{1}{1+\exp(-u_i^k)}$, and a value for each of the nodes in the next layer are calculated as $u_j^{k+1} = \sum_i w_{ij}^k g(u_i^k)$. The outputs are simply the values at the nodes in the final layer u_i^{l-1}, and are calculated from the inputs by progression through the network. The algorithm requires three sets of data: a training set, a validation set, and a testing set. The sets must contain information that we will write as $\underline{x} \rightarrow \underline{y}$ where $\underline{x}$ is the input vector and $\underline{y}$ is the output vector. (In the actual data this would just be a list $x_1 ... x_n, y_1 ... y_m$; but for clarity in discussion we will use a more abstract notation and make a distinction between inputs and outputs with $\rightarrow$.) The training set is the data used to develop the function $F_{\mathrm{ANN}}(\underline{x})$ by adjusting the weights w_{ij}^k to minimise some cost function. The cost function –

usually a simple least squares $C = \sum_k |F_{\rm ANN}(\underline{x}) - \underline{y}|^2$ – is minimised by an iterative process. Perhaps the most common method is 'back-propagation', which involves calculating the derivative of the cost function with respect to each of the weights from the last layer backwards, and using this to find the next estimate for the minimum until some convergence criterion is satisfied. A more detailed discussion can be found in [97]. A regularisation term of the form $R = \sum_{i,j,k} (w_{ij}^k)^2/\alpha$ may be added to the cost function to prevent any of the weights becoming too large. It is straightforward to see that the cost function is just the negative of a log-likelihood (up to an additive constant), assuming independent Gaussian likelihoods around the true values $\underline{y}$ with unit variance. (The cost function may be easily altered to provide different variances.) Adding a regulariser is equivalent to including Gaussian priors on the weights centred on zero. As such, minimising the cost function is equivalent to maximising some log-likelihood, although the ANN does not provide a posterior distribution.

In order to avoid over fitting to the training set, convergence is reached when the cost function for the validation set, rather than the training set, ceases to improve. (Both sets therefore need to be representative of the total sample.) This means that none of the data points in the validation set have actually been used as training data, which avoids the problem of simply "joining the dots" in machine learning and goes instead on its predictive power. These two sets are all that are used to fix the weights and hence the function $F_{\rm ANN}(\underline{x})$. After this we may use the testing set to check how well the ANN predicts the function we wish to model, and calculate the scatter around the line $F_{\rm ANN}(\underline{x}) = f(\underline{x})$. The test set does not necessarily have to be representative of the data. For example, if we wish to check how well the algorithm works for a subset of the population with extreme properties, but of which there are not enough to provide reliable training on their own, we can use this subset as a testing set (excluding any items used in the training or validation) and gain an understanding of how well the solution performs in different regions.

It is worth noting that the flexibility of ANN algorithms goes beyond their freedom from basis functions and linearity. It can be shown, as in [84], that an

arbitrary function of a (finite dimensional) vector $f(\underline{x}) : \mathbb{R}^n \to \mathbb{R}^m$, satisfying some basic smoothness/continuity conditions that one might reasonably expect from a physical system, can be modelled to an arbitrary degree of accuracy by a finite linear combination of compositions of sigmoid functions. This is very important for demonstrating the power of these algorithms, and gives us hope that if a (reasonably simple) relationship exists between the variable that we wish to model ($M_{\rm sim}$) and some additional parameter set $\underline{x}$ then we might be able to uncover it using ANN techniques, even if its form is complex and non-linear. The results from an ANN will vary depending on a number of factors, including the particular algorithm implemented (e.g. the method of finding the maximum likelihood or minimum cost), the structure of the network used (number of layers, and number of nodes in each layer), and the number (and combination) of input parameters (for instance, adding in lots of parameters which are not useful may affect the performance of the algorithm adversely). Despite the proof in [84], we must acknowledge some subtleties: we know that a suitable function may be modelled to a chosen degree of accuracy by combination of sigmoid functions using a *finite but unbounded* set of nodes. This means that although a solution exists, a solution will not necessarily exist with the node structure that one is using, and the number of necessary nodes cannot be known a priori. Furthermore, the proof is not concerned with finding the solution, only that it exists, and hence the algorithm is not guaranteed to converge on the correct solution even if it is within the scope of the network structure. Finally, there may be noise parameters to deal with, which will mask the smoothness of the function and invalidate the assumption that for a given input $\underline{x}$ there is a unique value of the function $f(\underline{x})$. Although we are left with no guarantees, ANN are a promising machine learning method which have had many successful applications, and their mathematical potential is evident. ANN are already being applied to cosmology, particularly in the field of photometric redshifts such as [82], [124]. A thorough, but accessible, introduction to the topic can be found in [113].

4.4.2 Machine Learning and Physics

It is worth pausing for a moment now to consider the implications of applying machine learning (ML) methods to a physical problem, particularly a method as complex as ANN, and how one ought to interpret such results. We have long applied ML to areas of physics such as photometric redshifts, where we are interested in results but not in the physics itself (see [82], [124] for example). In the case of photometric redshifts, or classification of celestial bodies (galaxies, supernovae), we are only interested in the empirical relationship between our inputs and outputs, as the physics is generally well understood; the problem is rather that we don't have the time or the information to analyse the vast numbers of objects. In the case of the LG mass however we are not only interested in obtaining a mass estimate, but also exploring some neglected physical aspects of a the problem. In this case we attempt not only to train the machine to do a job for us, but we also inspect the solution upon which it settles.

At its most fundamental, ML algorithms simply select from a class of functions that which best fits the data available to it. In the case of ANN, this class of functions is very large and powerful; [84] demonstrates that any suitably smooth function is, in principle, within the function class of ANN to an arbitrarily good approximation. We would certainly expect a problem such as classical orbital dynamics to produce smooth functions, so what is likely to be the limiting factor on an algorithm's success? The answer lies in the *input* to the function. Let us consider a physical problem where some quantity is a function of three parameters $\phi = F(x,y,z)$. If we train an ANN on, say $(x,y) \rightarrow \phi$, we have incomplete information, and the algorithm will search for the best function $G(x,y) \approx \phi$. What does the output then tell us? It shows us firstly how well we can do with the parameters (x,y), and secondly by plotting the function $G(x,y)$ we can see the projected dependence on our parameter subset. This is rather like marginalising over an unknown parameter, which is being done implicitly by analysing data without access it. In the case of the mass of the LG, we may have something like $M_{LG} = F(r, v_r, v_t, \delta\rho, ...)$, and by applying ANN we can see both whether or not particular parameters contribute

to the problem and its solution, and how sufficient we the parameter subset is. We might ask, for example, if (r, v_r) is enough to calculate reliable masses, despite the inaccuracy of $M_{TA}(r, v_r)$ is some regimes. In other words, is there a more accurate function of these two variables or is the TA the best we can do with incomplete information? If such a function exists, we expect the ANN to be able to find it. If other variables are significant, what does the solution function look like? Can we interpret this physically, and does it lead us to new analytic considerations? These are some of the questions we should bear in mind when applying ML methods.

Despite the complexity of the inner working of the ANN – there may be dozens of weights, and the composition of dozens of sigmoid functions is hardly trivial to understand – we have some avenues open to us in analysing the output function of the ANN. Once the function is derived we may apply any inputs we wish and check the output. This means we may plot cross-sections of the function, varying one or two variables whilst keeping others fixed, to observe how the output changes with respect to the inputs. (We shall do this in Figures 4.5a & 4.5b in section 5.3.2.) This potential to visualise the function and its dependencies may lead to physical insights if we can understand how these trends might come about.

4.4.3 Applying ANN to mass estimation

To apply the ANN to the problem at hand our observables such as r and v_r, divided by some appropriate unit of measurement, are used as input values i.e. the first layer of our network. The simulation masses are the target outputs used in the training and validation sets. In order to produce more stable calculations and avoid large weights (which lead to very large regularising terms), units are scaled so that all inputs and outputs are within a few orders of magnitude of unity, with the following units being used:

1. Mass in $M_G = 10^{12} M_\odot$
2. Time in Gyr $= 10^9$ years
3. Distance in Mpc
4. Velocity in Mpc Gyr^{-1}

The network needs only one node in the final layer for the output, which is the predicted mass in M_G. The ANN algorithms have randomised elements which require a random seed which can produced small variations in the results; because of this we use an ensemble of five ANN regressions with different seeds to produce each result.

All ANN runs were required to converge to be included in the ensembles. The architecture used for the results included in this paper had 5 hidden layers of 10 nodes each; more complex architectures did not yield an improvement indicating that the functions were not being limited by their number of free parameters.

The training, validation, and testing sets include ≈ 5000 halo pairs each, selected at random from the full set (subject to any data cuts) and with no halo pair appearing in more than one of the three sets. These contain, for each pair, the input parameters $\{r, v_r, ...\}$ and the target parameter $M_{\rm sim}$. Performance is tested by comparing the predicted values of the ANN on the testing set with the genuine values $M_{\rm sim}$ for each pair therein.

4.4.4 Comparing the ANN to the TA with Λ

We would like to get an idea of how much improvement the machine learning methods have offered compared to our earlier application of the TA. We can measure then the rms scatter of the predictions from their 'true' mass taken from the simulation. We apply the ANN directly to $\log(M_{\rm sim})$ using (r, v_r) as inputs, and compare this to the prediction from the TA.

To calculate the scatter we the use simple measure:

$$s_{rms} = \sqrt{\frac{\sum_i (\log(M^i_{\rm p}) - \log(M^i_{\rm sim}))^2}{N}} \tag{4.3}$$

Where M_p is the predicted mass (ANN or TA), $M_{\rm sim}$ is the mass from the simulation, and N is the number of points in the sample. This is the measure that is typically taken (plus an additional regulation term) as the cost function of the ANN and other machine learning algorithms.

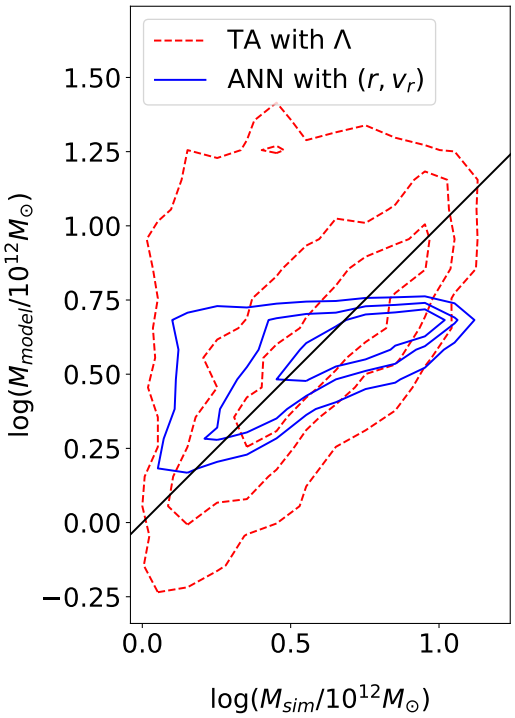
The results are summarised in Table 4.1 and Figure 4.2a. At a first glance, the

Model	rms scatter	Pearson correlation	Mean Bias
TA with Λ	0.43	0.32	0.345
ANN	0.23	0.53	-0.024

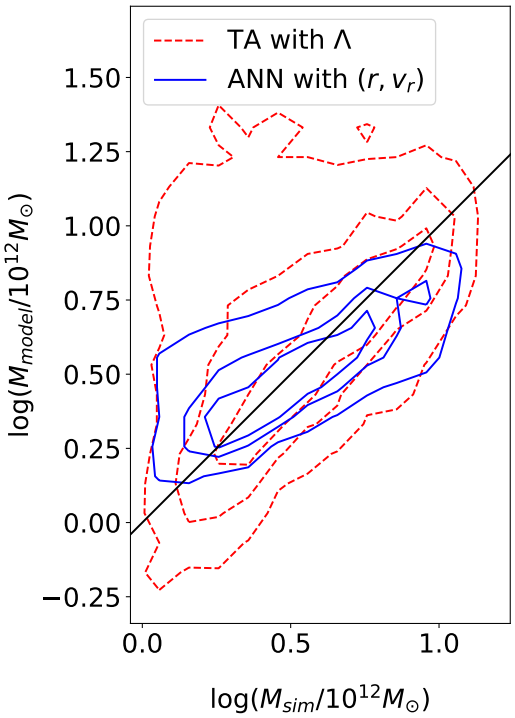
Table 4.1: Root mean square scatter in the log between predicted mass and simulation masses, using the TA or the ANN to predict the masses of the halo pairs using only the dynamical parameters (r, v_r). We also include the Pearson correlation coefficient, and the mean bias $\frac{1}{N}\sum_i(log(M_{\rm p}) - log(M_{\rm sim}))$. Despite the obvious failings on the ANN model in this case (see figure 4.2a), the summary statistics all show improvement. The ANN, though failing to find a reliable relationship, has nevertheless centred the output so that the solution performs well under standard metrics. Using machine learning with standard metrics as cost functions can easily lead to clearly problematic solutions if they are not properly inspected.

numerical results in Table 4.1 suggest that machine learning methods can reduce the scatter by almost half using only the same inputs as the analytic model, and significantly improve the correlation coefficient as well. This would imply that if we feel confident in our distribution of halo masses within large cosmological simulations, then machine learning methods may provide a significant improvement upon TA_Λ.

Figure 4.2a however demonstrates an interesting problem. The shape of the contours are not well in line with the equality line $M_p = M_{\rm sim}$, indicating that the estimate is highly biased, and the shape of the contours is notably compressed into the middle of the range. Considering that the TA is very successful for many halo pairs, and that it is easily reconstructable by an ANN, can we understand why the ANN produced the behaviour that it did? Despite the fact that the TA appears to show much less bias for many halos, there are a very large number of low mass galaxies which are drastically overestimated by the TA. These would contribute very heavily to the cost function of the ANN $\sim \sum(M_{\rm p} - M_{\rm sim})^2$ if it tried to reproduce something like the TA. With only (r, v_r) as inputs however, the ANN cannot find a solution which is satisfactory for these outlier halos, and therefore finds it beneficial to compress the estimates down into a narrower mass region clustered around the mean mass. This has happened to an extent that the function appears to have little visible correlation with the mass; high mass estimates in particular (where the simulation mass spans almost the entire range) are moved towards the centre of the mass range



(a) ANN, run with only (r, v_r) as input, and TA estimates of mass against the bound mass from simulations.



(b) Contours comparing the ANN model utilising shear with the mass in the simulation. The result shows significantly better correlation than ANN estimates without shear (correlation coefficient = 0.63).

Figure 4.2: Contours comparing the mass estimates from the ANN with those from the TA. TA (with a Cosmological Constant) is shown by the dashed red lines, the relevant ANN estimate in solid blue, with the diagonal (equality line) overplotted. Points here are taken from the 'testing' subset of the data (roughly 5000 points); contours are drawn as in Fig. 1, but due to the smaller sample size are rougher and enclose 30%, 60%, and 90% of the pairs. The equality line is over plotted in black.

to minimise the scatter. There is another potential problem which could cause behaviour like this. The sample is not uniform over the mass range, since it is subject both the halo mass function, which biases the sample towards the low masses, and the halo pair selection procedure, which shapes the mass distribution by selection MW-M31 analogues. This distribution of masses can create an effective prior on the output, since in lieu of a working model the best way to minimise the cost function is to send the output towards the densest region of the sample. The halo mass histogram peaks roughly in the middle of its log distribution, meaning that values

Table 4.2: Table showing the rms scatter in the log mass for each ANN model using different input parameters and cuts to transverse velocity, alongside the TA_Λ for comparison. For all ranges of v_t the ANN with input $(r, v_r, \lambda_i, \mu_i)$ reduces the scatter in the log significantly compared to the TA: by approximately 50% when there is no cut on v_t, or approximately 40% when there are strict cuts.

Max v_t	(r, v_r)	(r, v_r, v_t)	(r, v_r, ρ)	$(r, v_r, \lambda_i, \mu_i)$	TA_Λ
None	0.23	0.23	0.23	0.21	0.41
500 km/s	0.23	0.23	0.23	0.20	0.39
250 km/s	0.23	0.22	0.22	0.20	0.35
125 km/s	0.20	0.20	0.20	0.17	0.28
62.5 km/s	0.18	0.19	0.17	0.13	0.21

for the log-mass are pushed to the middle. These are a fairly generic problems in machine learning when presented insufficient information to solve a problem: the best way to reduce the scatter is to cluster around the mean. If the TA is to be improved upon, we will need to find out why some galaxies have much lower masses than we naïvely expect, or find a parameter which can break the degeneracy.

4.5 Extending the Model and Cuts to the Data

We now look at the effect of using some of the additional parameters we have at our disposal from the large simulation data set.

4.5.1 Transverse velocities

The first additional parameter used was the transverse velocity. We use the modulus of the transverse velocity as a one dimensional quantity rather than using a two dimensional vector, since the direction (which must be perpendicular to the radial vector) should not be relevant in the absence of other information. Thus $v_t = |\mathbf{v} - \mathbf{v_r}|$. We run the ANN with input $(r, v_r, v_t) \rightarrow \log(M_{\rm sim})$.

Adding the transverse velocity on its own has little effect on the scatter (see Table 4.2), where we show the scatter in the log for ANN results with different additional parameters and cuts.

4.5.2 Cuts to transverse velocity

Cuts to the transverse velocity have a significant impact on the scatter (see Table 4.2). Samples which are limited to small transverse velocities show significant im-

provement in both the TA and in the ANN. For the TA this is obvious – the argument is based upon the assumption of a radial orbit. The ANN improves alongside it, with the ANN scatter remaining $\sim 40-50\%$ less than the scatter for the TA for each cut. This may be because the range of possible orbits that the ANN has to reconstruct is restricted when the system is closer to one dimensional, leading to a simpler target function.

4.5.3 Environmental parameters

We may be able to improve our results by including parameters which characterise the environment of the halos. We are interested in observable parameters which may affect the dynamics of the pair, such as the large scale density field, velocity and shear fields.

Environmental parameters show that some improvement may be made compared to the application using only the traditional dynamical variables (distances, velocities).

4.5.3.1 Density

The density parameter $\rho/\bar{\rho}$ (smoothed over 2Mpc) was added to the ANN, although the improvement is not significant (see Table 4.2). Perhaps a more fruitful parameter would be a density gradient vector, which would give an indication of the strength and direction of the external gravitational field which may perturb the orbit, although this is beyond the scope of this work.

4.5.3.2 The velocity shear tensor

The most useful parameter investigated in this chapter is the shear tensor (also smoothed over 2Mpc). The shear tensor ([100], [91]) is defined as

$$\Sigma_{ij} = -\frac{1}{2H_0}\left(\frac{\partial v_i}{\partial r_j} + \frac{\partial v_j}{\partial r_i}\right) \tag{4.4}$$

At a given point the eigenvalues and eigenvectors of this tensor can be calculated. The shear tensor has three eigenvalues $\lambda_{1,2,3}$ with three unit eigenvectors $\hat{e}_{1,2,3}$. We

order the eigenvalues such that:

$$\lambda_3 \leq \lambda_2 \leq \lambda_1 \tag{4.5}$$

We calculate the absolute value of the cosine of the angle between the eigenvectors and the radial separation vector $\mu_i = |\cos(\theta_i)| = |\hat{e}_i \cdot \hat{r}|$. (We must use the absolute value since the eigenvector only defines an axis rather than a specific direction, so alignment with $\hat{e}_i$ or $-\hat{e}_i$ is physically identical.) We use the ANN with an input of the form $(r, v_r, \lambda_1, \mu_1, \lambda_2, \mu_2, \lambda_3, \mu_3) \rightarrow \log(M)$. This provides the ANN with information about how the orientation of our physical system lines up with the principles axes of any local expansion or collapse. The scatter in the log mass, as shown in Table 4.2, reduces by about 10–15% compared to the most basic form using (r, v_r) with an equivalent cut on v_t. Perhaps more importantly, there is a change in shape of the contours between Figure 4.2a (the results without environmental information) and Figure 4.2b (the results utilising shear information). The upper bound on the estimated masses has now been relaxed and the results look significantly more intuitive, with a less dramatic bias. We can see from Figure 4.4 that the tail TA_Λ overestimates is significantly reduced. Nevertheless, the ANN with shear information still produces a noticeably biased relation, which underperforms compared to the inner contours of the analytic TA_Λ. This is because there are still a large number of pairs in the full sample which the ANN cannot solve for given the information that it has, even though it is able to make an improvement. It is interesting to note that when the data is cut for low v_t, the bias in the ANN model greatly reduces (fig 4.3).

Why might the inclusion of shear have improved things? The velocity shear tensor contains information about the movements of matter on large scales, and hence may be linked to variations in behaviour from what one would expect in an isolated environment. The tensor may be used as a map of the cosmic web (as in [100], [91]), with the sign and magnitude of eigenvalues determining if a halo is occupying a knot, filament, sheet, or void. Such a classification method has already been used to examine how halos and galaxies align with the large scale

Figure 4.3: Contours comparing the mass estimates from the ANN utilising shear information with a sample from the full data set (blue, dashed) and a sample with low $v_t < 62.5\,\mathrm{km}\,s^{-1}$ (red), with the diagonal (equality line) overplotted. Contours enclose 20%, 50%, and 85% of the pairs. The equality line is over plotted in black.

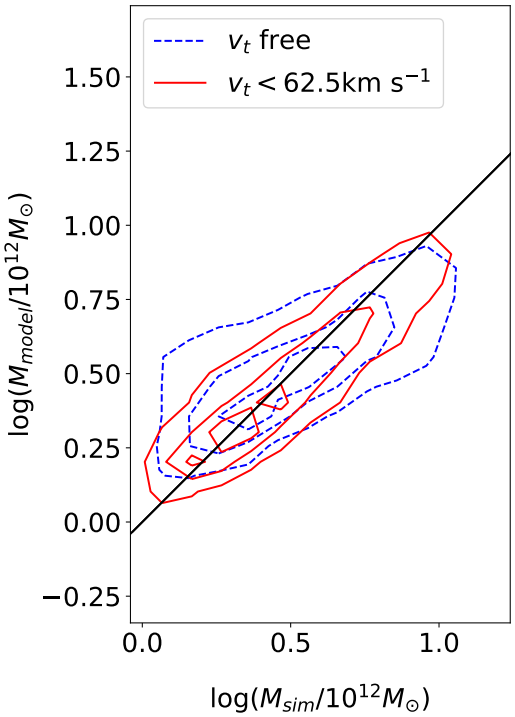
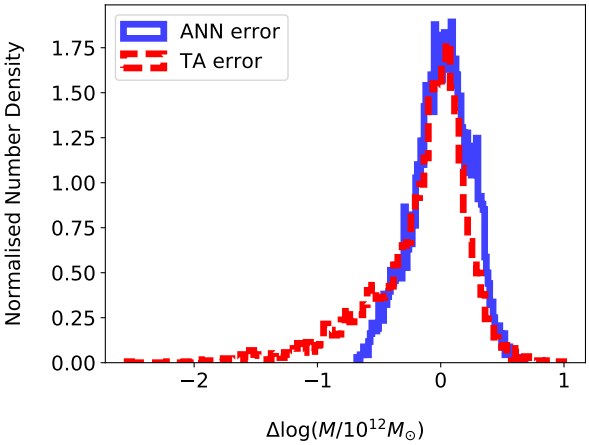
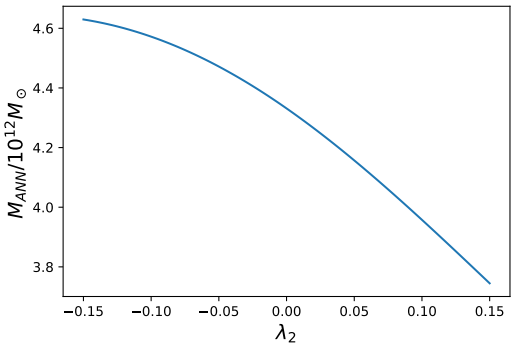
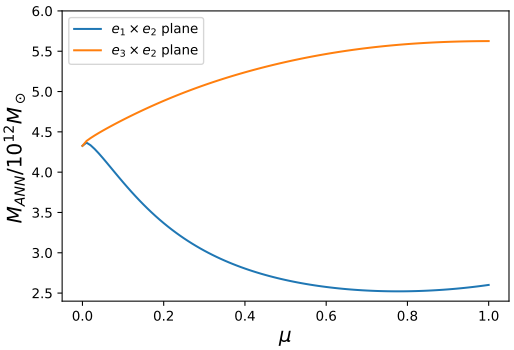


Figure 4.4: Error histograms for the TA and the ANN run with shear information (with no cuts on v_t). The ANN manages to avoid the large tail of overestimates that is produced by the ANN. $\Delta\log(M) = \log(M_{\mathrm{sim}}) - \log(M_{\mathrm{model}})$.





(a) The dependence of ANN mass estimate on the parameter λ_2 using $\mu_2 = 1$ (i.e. the system is totally aligned with the second eigenvector e_2), $\mu_{2,3} = 0$. To respect the ordering of $\lambda_{1,2,3}$ that we have used, the upper and lower eigenvalues are set at $\lambda_1 = 0.15$ and $\lambda_3 = -0.15$. Negative λ implies expansion and positive λ implies collapse along the direction $\hat{\mathbf{e}}_\mathbf{i}$. Masses estimates are higher in expanding regions and lower in collapsing regions.



(b) The dependence of the ANN mass estimate on the alignment with the eigenvectors. Here we have held the eigenvalues at $\lambda_1 = 0.15$, $\lambda_2 = 0$, and $\lambda_3 = -0.15$ i.e. collapse along the e_1 direction, no shear on the e_2 direction, and expansion along the e_3 direction. The angle is measured between the radial separation vector and the eigenvector e_2, when rotating in the $e_3 \times e_2$ plane (from full alignment with e_2 to full alignment with e_3), and rotating in the $e_1 \times e_2$ plane (from full alignment with e_2 to full alignment with e_1). We see that as we rotate to align with the collapsing vector (e_1) our mass estimate decreases, and as we rotate to align with the expanding vector (e_3) our mass estimate increases.

Figure 4.5: Cross sections of the ANN function $F_{ANN} : (r, v_r, \lambda_i, \mu_i) \rightarrow \log(M)$ where one parameter is varied and the others kept constant to visualise the mass dependence on that parameter for a LG-like halo pair. In both cases $r = 0.77$ Mpc, and $v_r = -130$ km s^{-1}, and these parameters are not varied.

structure ([102], Forero-Romero, Contreras, & Padilla 2014, [116]), how galaxy properties are affected by their cosmic web environment ([115], [114]), and how LG-like systems are formed in filaments ([88]).

In order to understand something about the way that the shear is influencing the solution, we look at the output with some physical insight and intuition. Whilst the fundamental representation of the function – encoded in dozens of weights – is no doubt difficult to interpret on its own (it is, after all, a rather garbled way to write a function and even the most brilliant mathematician would balk at diving into a quagmire of sigmoid functions analytically), we may achieve significant insight by plotting cross sections of this multidimensional function. By varying just one variable at a time and holding all others constant we may see the impact of that variable

on systems which are otherwise totally similar. For example, we may hold (r, v_r) fixed, thus considering only systems which have identical observable dynamics at present, and vary just one of the shear magnitudes λ_2. (We choose λ_2 since we must always respect $\lambda_3 \leq \lambda_2 \leq \lambda_1$ which can be done by setting $\lambda_{1,3}$ and varying λ_2 between them.) This is shown in Figure 4.5a. We wish to inspect a case where λ_2 has great impact and the other shear parameters have as little as possible, so we choose a system which is aligned totally with e_2. The result is the mass estimate decreases with increasing λ_2, i.e. with increasing degree of 'collapse' along the radial axis of the system. This is the behaviour of the function that has been derived, and now it is prudent to ask whether this makes physical sense. In order to ask such a question, we must be very sure about what has changed and what has not. The background has changed, but the observed dynamics have not. Let's compare to a case with no shearing at all – two bodies orbit one another in empty space. Heavy halos turn around sooner and will fall into one another faster; lighter halos will turn around later and fall into one another more slowly. Let's say in the empty space scenario, a mass of M_* will yield the dynamics we see (r, v_r). Now let us consider the case where there is a larger scale tendency towards compression along the radial axis (high λ_2). Our smaller bodies are also part of this flow, and will be involved in this larger scale collapse. This will bring us to a quicker turnaround than the empty space case, and thus we will not produce the same dynamics with M_*. In order to delay our turnaround and slow the collapse back down to our observed dynamics, we must reduce the mass estimate, and so we have a collapsing mass $M_C < M_*$. The same reasoning may be applied to the expanding case to yield an expanding mass $M_E > M_*$.

It is worth at this point addressing some issues of structure formation. We may look towards the Raychaudhuri equation, or its Newtonian equivalent, as a means of calculating the mass in the presence of shear. Here, crucially, the shear only enters as $\sigma^2 = \sigma_{ij}\sigma^{ij}$, and thus any shear regardless of sign or alignment (the case is spherical) will enter in the same way; when applied to structure formation we obtain a reduced mass. As a result, one might expect that we should see our mass estimate

reduce in the presence of any shearing, but this is not the case. Remember that we are not tracking the growth of an initial over density, but looking at the dynamics of two bodies and asserting the mass which would produce it. The presence of shear might cause the galaxies to initially form with lower masses than they would otherwise, but the mass with which they are formed is irrelevant to our purposes: we only see deviations of the trajectories from empty space dynamics, e.g. when light galaxies look heavier than they are due to their distorted dynamics. This is very similar to why the density parameter is not useful – it affects the formation of halos but, if it is not much varying over the distance scale of the orbit, it does not affect much the dynamics of the halos.

Figure 4.5b explores the significance of the alignment of the radial vector with the shear eigenvectors. The eigenvalues are set up so that one axis is 'collapsing', one axis is 'expanding', and the other is 'neutral' ($\lambda = 0$). The observed dynamics (r, v_r) are again kept constant, and the alignment of the radial vector is rotated (via the parameters $\zeta_{1,2,3}$) from total alignment along one eigenvector to the next. The results show that the impact of the collapsing and expanding effects that we previously discussed are dependent on the alignment of these axes with the radial axis. If there is collapse close to orthogonal to the radial system, it does not much impact the trajectories along the radial line, so there is little change to mass. This is physically to be expected.

4.5.3.3 Shortcomings of additional parameters

The scatter is still very significant, and the additional parameters make a relatively small impact on the result. Our inability to reduce it further may be for a number of reasons

1. There are further important variables which may have a more significant effect.

2. The relationship between the dynamics of the pair and the environment is highly complex and non-linear, perhaps even chaotic.

3. We only have a snapshot of these parameters at the present time; they are

Table 4.3: A table of observed velocity values for the LG from three papers. van der Marel & Guhathakurta (2012) is the most reliable recent result. van der Marel & Guhathakurta (2008) is included for comparison with earlier papers such as Partridge et al. 2015, which use values for the LG closer to this. Salomon et al. (2015) is a controversial recent result suggesting very different relative motion of MW and M31, which is included to explore the implications of such a result. These parameters will be referred to in the text and table as vdM. 2008, vdM. 2012, and Sal. 2015 respectively.

	vdM. 2008	vdM. 2012	Sal. 2015
v_r / km s^{-1}	-130 ± 8	-109.4 ± 4.4	-87.5 ± 13.8
v_t / km s^{-1}	42 ± 56	17 ± 17	149.4 ± 92.3

Table 4.4: A table presenting the mass estimates for the local group for each model, as applied to the input parameters with velocities from vdM. 2008, vdM. 2012, and Sal. 2015 (see Table 3). Quoted errors are first for propagating the errors in the observables through the TA/ANN functions, and second for $\pm$ the r.m.s. scatter on the log mass between the TA/ANN model and the simulation set. For vdM. 2008 and vdM. 2012 we may use strict cuts on v_t, whereas for Sal. 2015 we must use cuts that include the higher transverse velocity, thus leading to a higher uncertainty. The Bayesian results quote only a 68% confidence interval.

	M_{LG} / $10^{12} M_\odot$		
Model	(vdM. 2008)	(vdM. 2012)	(Sal. 2015)
TA_Λ	$5.8^{+1.0+3.6}_{-0.9-2.2}$	$4.7^{+0.7+2.9}_{-0.6-1.8}$	$3.8^{+1.1+4.7}_{-0.9-2.1}$
ANN	$3.7^{+0.3+1.9}_{-0.3-1.3}$	$3.6^{+0.3+1.8}_{-0.3-1.2}$	$3.3^{+0.6+2.3}_{-0.5-1.4}$
ANN + Shear	$6.1^{+1.1+2.1}_{-1.1-1.6}$	$\mathbf{4.9^{+0.8+1.7}_{-0.8-1.3}}$	$3.6^{+1.3+2.1}_{-1.1-1.3}$
Bayesian	$3.4^{+1.9}_{-1.2}$	$3.1^{+1.3}_{-1.0}$	$3.4^{+2.3}_{-1.3}$

however dynamical and the impact on dynamics would depend on the variation in these parameters throughout the history of the universe. Although this is something we can extract from simulations, this is not observable in the actual LG, and thus isn't something we can take advantage of for calculations.

4.6 Application to the Local Group

We may apply what we have learned directly to the LG i.e. the Milky Way (MW) and its most massive partner, Andromeda (M31). We have learned that the TA's range of applicability may be narrow, and both the TA and ANN are more successful

if we make cuts to look at LG-like galaxy pairs. Results of the different methods of calculating the LG mass is summarised in Table 4.4.

4.6.1 Input parameters

We use $r = 770 \pm 40$ Mpc, and velocity parameters from three studies referred to as vdM. 2008, vdM. 2012, and Sal. 2015 (see Table 3 for details). The most reliable case is represented by vdM. 2012, from [126]. vdM. 2008 is taken from an earlier paper and included only in Table 4.4 for comparison to earlier results. [125] (Sal. 2015) represents a controversial new study which suggests a non traditional relative motion, which a much larger transverse motion. For both cases we will use the same range of r, as it contains the values of r used in both papers.

The shear parameters of the local universe are taken from [101], and are $\lambda_1 = 0.148 \pm 0.038$, $\lambda_2 = 0.051 \pm 0.039$, and $\lambda_3 = -0.160 \pm 0.033$. To calculate the angle between the eigenvectors and the MW-M31 separation vector, we use [101] for the shear eigenvectors and [85] for the direction of the separation vector.

4.6.2 TA estimates

Our most accurate TA model is the TA with Λ, which applied to the LG with vdM. 2012 gives a mass estimate of $4.7^{+0.7}_{-0.5} \times 10^{12} M_\odot$ by propagating the errors in the input through the TA model. On top of this we must take into account the intrinsic 'cosmic scatter' which we can calculate from the r.m.s. scatter when we apply the TA to our data set (see Table 4.2). Applying this error on top of the upper and lower limits from the TA gives a LG mass of $4.7^{+2.9}_{-1.8} \times 10^{12} M_\odot$. Using the velocity observations with Sal. 2015, we have an estimate with propagated errors of $3.8^{+1.1}_{-0.9} \times 10^{12} M_\odot$, on top of which we have cosmic scatter of $3.8^{+4.7}_{-2.1} \times 10^{12} M_\odot$.

The primary reason for evaluating the TA_Λ mass estimate is to set a benchmark for the other methods, although it remains a significant estimator in the literature. The result we retrieve is slightly higher than the TA mass in [126], as we have included a Λ term.

4.6.3 ANN estimates

To calculate the estimates for each case, we use the ANN which we trained previously with cuts on transverse velocity. For the high transverse velocity case we may use $v_t \leq 250\text{km s}^{-1}$, and for the low transverse velocity case we may use $v_t \leq 62.5\text{km s}^{-1}$. One could approach this by making stricter cuts on each of the parameters to create a LG like set, but we do not have a large enough set to collect enough pairs for the training, validation, and testing sets in such a narrow band of parameters. Errors in the input parameters are propagated through by running the ANN with all variations of mean, max, and min parameters for (r, v_r, v_t).

4.6.3.1 Using only (r, v)

Using vdM. 2012, which has the more conventional (smaller) transverse velocity, we have an estimate of $3.6^{+0.3}_{-0.3} \times 10^{12} M_\odot$ by propagating errors from the input. This error is dwarfed however by the random scatter, which gives us $3.6^{+1.9}_{-1.3} \times 10^{12} M_\odot$. With the higher transverse velocity and lower radial velocity of Sal. 2015 we get an estimate of $3.3^{+0.6}_{-0.5} \times 10^{12} M_\odot$ by propagating the errors through. Once again, this is small compared the intrinsic scatter, $3.3^{+2.3}_{-1.4} \times 10^{12} M_\odot$. The very small variation due to propagating the uncertainties in the inputs is a consequence of the clustering around the mean mass; the ANN trained on (r, v_r) is strikingly insensitive to input.

4.6.3.2 Including shear

The estimate of the LG mass is boosted when shear information is taken into account. Using this ANN model, we obtain estimates of $4.9^{+0.8+1.7}_{-0.8-1.3}$ and $3.6^{+1.3+2.1}_{-1.1-1.3}$ for vdM. 2012 and Sal. 2015 respectively (propagated errors are quoted first, followed by intrinsic scatter).There is a somewhat larger propagated error from the uncertainties in the input than even the TA_Λ; this is both due to having more parameters in our inputs and the difficulty in acquiring very accurate shear information. The velocity shear tensor identifies the MW-M31 system as being in a filament; the largest collapse eigenvector is almost perpendicular to the radial separation vector (and hence can be safely ignored), but there is an appreciable component along the other two. Although one of the eigenvalues indicates collapse, we also have a large

expansion term with an appreciable component along the radial separation direction. This may be causing the overall boost in the system mass when included in the calculation. The effect of the shear compared to the TA is well within the errors of either method, so it is not necessarily possible to say what additional dynamics are at play, or whether the mass estimate is distinctly higher in the presence of shear for the case of the LG.

4.6.4 A probabilistic approach

Another approach is to look at the simulation data from a Bayesian perspective to estimate a probability distribution for mass, given the masses in the simulations. This can be calculated using an MCMC approach or a numerical integration. A simple approach is to start by cutting the data to select only binaries which have parameters within the limits set by observations of r, v_r, and v_t for the LG. A probability distribution for the simulation masses can then be fitted to this subset of the data, which is assumed to be Gaussian in the log mass with some mean μ and standard deviation σ. A probability distribution over the mass given the ensemble of masses from the simulation is given by the following expression:

$$P(M|\mathbf{M}_{\rm sim}) = \int P(M|\mu,\sigma,\mathbf{M}_{\rm sim})P(\mu,\sigma|\mathbf{M}_{\rm sim})d\mu d\sigma \tag{4.6}$$

where $\mathbf{M}_{\rm sim}$ is the data vector, composed of each binary i with mass $M^i_{\rm sim}$ in the cut of data (these data vectors are represented as histograms in Figure 4.6), and μ,σ characterise the probability distribution in the log mass. Since we don't know the mean and standard deviation of the underlying probability distribution a priori, we are marginalising over values of μ and σ in equation 4.6. We must then calculate the probability distribution on (μ,σ) as follows:

$$P(\mu,\sigma|\mathbf{M}_{\rm sim}) = \frac{P(\mathbf{M}_{\rm sim}|\mu,\sigma)P(\mu,\sigma)}{P(\mathbf{M}_{\rm sim})} \tag{4.7}$$

We must now make some assumptions in order to proceed. As mentioned above, we will assume that masses in $\mathbf{M}_{\rm sim}$, as well as the LG mass, are drawn from an

underlying Gaussian distribution in the log mass, and that samples are independent:

$$P(\log(M)|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(\log(M)-\mu)^2}{2\sigma^2}\right) \tag{4.8}$$

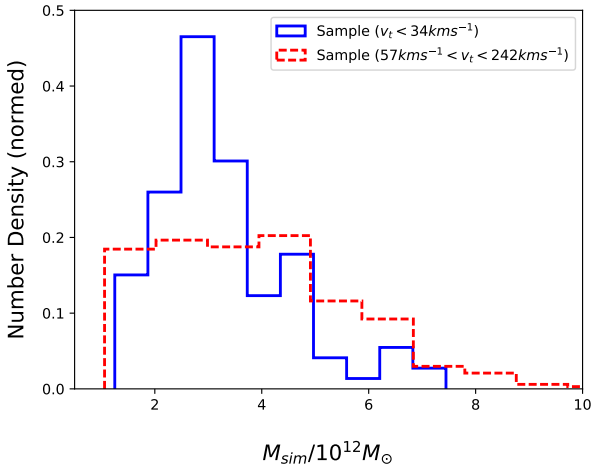
$$P(\mathbf{M}_{\rm sim}|\mu,\sigma) = \prod_i P(\log(M^i_{\rm sim})|\mu,\sigma) \tag{4.9}$$

We shall also assume flat priors (since we have no prior knowledge) on the parameters μ and σ, which means that both $P(\mu,\sigma)$ and $P(\mathbf{M}_{\rm sim})$ are simply multiplicative constants. (In practice we need to assume some hard limits on μ and σ so that we can terminate the integration, so their distribution is a top hat function which determines the integration limits for equation 4.6. The limits must be wide enough so that $P(\mathbf{M}_{\rm sim}|\mu,\sigma)$ is negligible outside; we have taken $\mu \in [0,1]$ and $\sigma \in [0.1,1]$.)

With these equations it is simple to evaluate $P(M|\mathbf{M}_{\rm sim})$ for a particular cut (see Figure 4.7). The posterior resembles to some extent the histograms in Figure 4.6, but the Bayesian approach allows us to take into account uncertainties – particularly those associated with the small number of samples – in a more robust way, as well as allowing confidence intervals to be calculated. Cuts are made to match the error bars given at the beginning of the section with the exception of v_r which ranges between 0.5 and 1.5 times the best estimate of v_r in each case in order to give use enough pairs. (A strict cut on all parameters yields less than two dozen pairs for each set.) For a system with low transverse velocity we find a mass of $3.1^{+1.3}_{-1.0} \times 10^{12} M_\odot$, and for high transverse velocity we find a mass of $3.2^{+2.1}_{-1.2} \times 10^{12} M_\odot$ at the 68% confidence level. Best estimates are calculated by the median because this is the same point whether we use a distribution in M or log(M), whereas the peak moves. The median is the peak of the distribution in the log, which is most comparable to the other results since they were calculated using in the log mass. This is largely compatible with the results from ANN, although errors have been calculated in a slightly different way so they may not be directly comparable.

The disadvantages of this method for generating a precise estimate are twofold. Firstly, we do not have enough pairs to sufficiently cut the data to match observations whilst leaving sufficient pairs to make a statistical judgement calculable.

Figure 4.6: Normalised histograms of the simulation samples, with each estimate of transverse velocity. There are 118 galaxy pairs with $v_t \leq 34$ km s^{-1}, and 349 with 57 km s^{-1} $\leq v_t \leq$ 242 km s^{-1}.



(Even with the relaxed cuts we have made, there are only $\mathcal{O}(10^2)$ pairs, which is not enough to make a strong statistical statement on such a varied group.) Secondly, it does not produce any information on the dependence on different parameters. In order to do this, we would have to create a parameterised probability distribution $P(r, v_r, ...)$ and solve for it. It is challenging to produce a flexible enough function form for this distribution in terms of multiple inputs without essentially entering the world of ML. (Methods such as ANN can, after all, be cast as a maximal likelihood method.) Rather, this procedure may be taken as illustrative, and providing a rough probability distribution of LG-like galaxy masses. It demonstrates that the distribution is broad, particularly when the transverse velocity is higher, and the varied results in the literature (see Table 4.5) are well within a sensible range of this distribution.

4.7 Discussion

We find that, despite some advance in the analytic models, the application of ML provides a rapid route to improved estimates of the LG mass. Analytically, the inclusion of Λ in the TA can provide an important improvement in the bias of the mass estimate when analysing a large collection of galaxies from ΛCDM simulations, however the inclusion of v_t is if anything disruptive. The uncertainty in *which*

Figure 4.7: Normalised posterior distributions for the mass of the LG calculated from equation 15, with each estimate of transverse velocity.

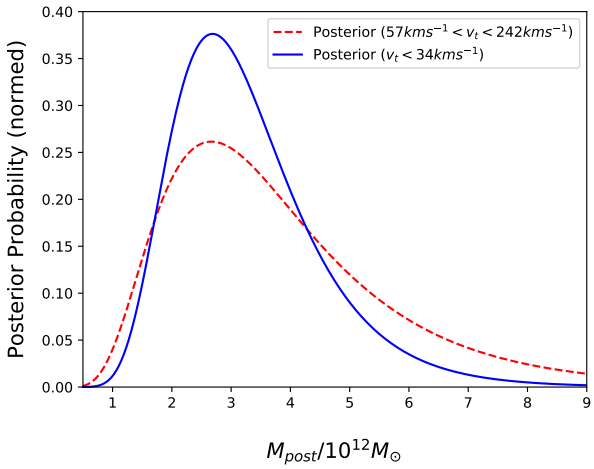


Table 4.5: A table comparing LG mass estimates of some recent papers. Errors between papers have been estimated with various methods and confidences, and may not be directly comparable. Results are compiled from Li & White (2008) [98]; Phelps, Nusser, & Desjacques (2013) [120]; Peñarrubia et al. (2015) [118]; González, Kravtsov, & Gnedin (2014) [90]; van der Marel (2012) [126]; Carlesi et al. (2016) [81].

Paper	Method	Mass	Comments / other Esimates
[98]	Large ΛCDM n-body Sim.	$5.27^{+4.96}_{-1.82} \times 10^{12} M_\odot$	$r = 784$ kpc $v_r = 130$ km s^{-1} $M_{TA} = 5.32 \pm 0.48 \times 10^{12} M_\odot$
[120]	Numerical Action Method	$6 \pm 1 \times 10^{12} M_\odot$	$r = 790$ kpc, $v_r = 119$ km s^{-1} $M_{\rm MW} = 2.5 \pm 1.5 \times 10^{12} M_\odot$ $M_{\rm M31} = 3.5 \pm 1 \times 10^{12} M_\odot$
[118]	Extended LG Simulation	$2.3 \pm 0.7 \times 10^{12} M_\odot$	$r = 783$ kpc, $v_h = 300$ km s^{-1} Uses heliocentric coordinates
[90]	Large ΛCDM n-body Sim.	$4.2^{+3.4}_{-2.0} \times 10^{12} M_\odot$	$r = 770$ kpc, $v_r = 109.3$ km s^{-1} Finds $M_{MW} + M_{M31} = 2.4 \times 10^{12} M_\odot$
[126]	TA ($\Lambda = 0$) converted to virial	$4.93 \pm 1.63 \times 10^{12} M_\odot$	$r = 770$ kpc, $v_r = 109.3$ km s^{-1} $M_{TA} = 4.27 \times 10^{12} M_\odot$ Multi-probe $M = 3.17 \times 10^{12} M_\odot$
[81]	Constrained Simulations	$1.79 \pm 0.51 \times 10^{12} M_\odot$	0.35 Mpc $< r <$ 0.7 Mpc -135 km s$^{-1} < v_r < -80$ km s^{-1}

parameters to base our model around is a significant stumbling block for this calculation, and ML methods are a valuable tool in analysing the potential of different parameters to improve the solution. Furthermore we find that the application of ANN presents a more significant improvement than analytic extensions to the TA. When trained (on a suitably large population) using only parameters (r, v_r) one finds the scatter in the log mass which is reduced by approximately half, but the estimator highly biased towards the mean mass of the sample, so that it fails to make predictions. Additional parameters, particularly the shear field, pushes this scatter lower, approximately half of the TA scatter and with a significantly reduced bias compared to the ANN with only (r, v_r) when tested on a sample with no cuts to v_t.

The velocity shear tensor provides a key improvement to the ANN prediction of mass. Without it, the ANN cannot find an explanation for why many halo pairs which appear to have high dynamical masses (judging from the TA) actually have much lower masses in the simulations; this leads to the ANN driving the outputs towards the mean of the input sample in order to minimise the scatter. This behaviour is undesirable, and characteristic of problems which lack information. The addition of shear information breaks to some extent this uncertainty and produces an improved correlation between $\log(M)_{\rm ANN}$ and $\log(M_{\rm sim})$. The resul is still biased for the full sample, but bias reduces significantly as the data is cut for lower v_t and the dynamics become simpler to model. Since the influence of the shear is poorly understood from an analytic perspective, an investigation into the behaviour of the derived function from the ANN trained on shear information may be used to gain physical insight. A possible explanation is that there is a change in mass estimate due to the bulk motion on large scales affecting the motion of the binary system. In other words, our halo pairs cannot be taken to interact in isolation, but their place within in the cosmic web structure determines their motion to some extent due to local velocity flows creating 'expansion' or 'collapse' along different axes. This is an effect not previously included in studies of the LG mass, and hints at the deficiencies of our isolated physical models.

Applications of trained ANN using the observed LG parameters provide an

estimate of the mass of the MW and Andromeda (see Table 4.4). This is consistent with other mass estimates, of which some recent examples are summarised in Table 4.5. Some mass estimates of the LG are also summarised in [83], which tend to range from $2.5 - 5 \times 10^{12} M_{\odot}$. There is some tension in the results of the LG mass, with a divide between low estimates [118] [81], and high estimates [98][117][120][126]. One potential source of discrepancies is the assumption that $M_{\rm LG} \approx M_{\rm MW} + M_{\rm M31}$. In [90], they find a significant mass difference between the LG ($4.2 \times 10^{12} M_{\odot}$) and the sum of the two major galaxies ($2.4 \times 10^{12} M_{\odot}$). Part of this may be due to the fact that dynamical methods that work on the basis of a binary system implicitly include the mass of any bound satellites with the mass of the galaxy. Hence, the Milky Way or Andromeda mass found by such a method would include all the satellites that move with it, and thus push the mass estimates of the individual galaxies much higher. In [81] they find a particularly low LG mass, at only $1.8 \times 10^{12} M_{\odot}$. It is notable however that the range of r is much lower than in the other studies, being contained in the range 0.35 Mpc $< r <$ 0.7 Mpc, whereas others (including this one) use $r \sim 0.77$ Mpc. This may go some way towards explaining the comparably low result. For example, taking a TA estimate for the LG mass from the middle of their range, which would have $r = 0.525$ Mpc, and $v_r = 107.5$ km s^{-1}, gives $M_{TA_{\Lambda}} = 2.3^{+1.6}_{-1.1} \times 10^{12} M_{\odot}$, with the lower and upper bounds calculated using $r = 0.35$ Mpc and 0.7 Mpc respectively. The mean (using $r = 0.525$ Mpc) is just at the upper estimate of [81] ($1.8 \pm 0.5 \times 10^{12} M_{\odot}$), and with significant overlap in errors in the TA are taken into account. This suggests that the results in [81] may indeed be consistent with this paper (and others in line with the TA) after all.

Given that there is only one LG, statistical methods cannot overcome the possibility that the LG is a 'special case', occupying an less than maximally likely part of the distribution, due to some hitherto unknown or unobservable parameters. The action of additional bodies, as in [118] and [120], may have an effect as well, which could make the system highly sensitive to configurations. It is also suggested in [118] that the uncertainties in the sun's transverse motion may be a significant

source of error in determining correctly the relative motions; it also however asserts that the TA must be systematically biased by a factor of about 1.5 in order to match with their result, which is not seen in this study or in [98] for galaxies with small transverse velocities. All of these mass estimation methods have significant error bars, and overlap at the one or two sigma level. Refinement of these methods will be crucial to identify a genuine tension between these results, which could point to the need to new physical models.

When we apply the ANN trained on shear information, we have a mass estimate of $4.9^{+0.8+1.7}_{-0.8-1.3} \times 10^{12} M_{\odot}$, compared to a TA_{Λ} estimate of $4.7^{+0.7+2.9}_{-0.6-1.8} \times 10^{12} M_{\odot}$. These are very similar, although the scatter is significantly improved. Bayesian methods based on a cut of data from the simulation also offer a competitive estimate of $3.1^{+1.3}_{-1.0} \times 10^{12} M_{\odot}$, although this is forced to be use few parameters and thus effectively averages over any other influences such as shear effects. (Making strict cuts on the all the parameters, including shear parameters, would require a larger dataset than used in this paper; a Bayesian analysis based on cuts on all parameters would require millions of halo pairs.) It is therefore unsurprising that the Bayesian analysis produces a result similar to the ANN trained on just two parameters (see Table 4.4). It is interesting to compare the ANN results with the Bayesian analysis on the data; if we look at the distributions in Figure 4.7 we notice that the value estimated by the shear ANN is not by any means surprisingly unlikely. The distributions in Figure 4.7 really represent the distributions when shear effects are averaged over, and hence the probability of the ANN mass estimate in the context of the marginalised probability distribution is related to the probability of being in an LG-like local shear environment in the simulation.

To improve the estimates for the LG, one may wish to refine this estimate by using a larger simulation, selecting more strictly "LG like" pairs, and using additional criteria such as morphology as in [98]. Applications to the LG may also be complicated by uncertainties in additional parameters. The transverse velocity (usually taken to be small) has been contested recently, implying that strict cuts on v_t in LG-like sets may not be justified, which jeopardises the use of simple dynam-

ical models such as the TA. In such a case it is particularly useful to have methods such as ML; the ANN estimates have only slightly higher scatter for analyses with no cut to v_t as for sets restricted to low v_t. The results of both the exploration of the TA with a Cosmological Constant term (TA_Λ) and the significance of the shear tensor (Σ_{ij}) indicate that successful dynamical estimates of mass are unlikely to be independent of cosmological context. We note that taking into account both the cosmological model (in order to model expansion terms such as Λ or more complex models such as modified gravity or dark energy) and the larger scale structure (as traced by Σ_{ij}) is crucial to understanding how galaxies are interacting. Our methods are dependent upon ΛCDM simulations, and thus are only applicable within this framework. Estimates of the LG mass in the context of other cosmological models – alternative dark energy models, modified gravities, or warm dark matter – could yield different results and must be explored separately using appropriate simulations. It is possible that a greater understanding of galaxy formation physics will inform such mass estimates even better, providing clues to possible boundary conditions for dynamical arguments, or linking galaxy masses to their environments and cosmological models in a more direct way.

Chapter 5

The Local Group in the Dark Universe

Poor wandering one,
Though thou hast surely strayed,
Take heart of grace,
Thy steps retrace!

Mabel Stanley

5.1 Introduction

Having, in the previous chapter, looked at the dynamical mass of the Local Group (LG) in the context of local structure, we now address another facet of the problem: that dynamical mass estimates of the Local Group (LG) are cosmology dependent. When one restricts oneself to the simple basis of the Timing Argument (TA) as an isolated, one dimensional system, one finds that the mass estimate is dependent on the observed dynamics (distance and velocity), and on the cosmology through the strength of dark energy (modifying the acceleration equation) and the age of the universe (modifying the boundary conditions). The age of the universe is determined by more fundamental cosmological parameters such as the matter density, the density of dark energy, and the Hubble parameter. The tension present between estimates of H_0 from the CMB, galaxy velocity flows, and LSS [76][75] introduces uncertainty in the age of the universe, and thus in the TA mass estimate. Our ma-

chine learning driven approach in the previous chapter is unable to deal with these cosmological dependencies, since there is only one realisation with fixed properties. It therefore does not take into account the uncertainty in the cosmology, and an estimate based on a different set of cosmological parameters or theory of dark energy would require retraining with an appropriately set up simulation.

We examine a variety of DE models and calculate the LG mass dependence on their parameters in the case of a simple, one dimension TA. One can infer a cosmologically marginalised LG mass by using the posterior probability distributions from numerical sampling and cosmological data; the LG mass may in future be determined from joint observations of the LG dynamics, tidal streams, large scale structure, and even the CMB! Conversely, in principle at least, if one had a good independent estimate of the LG mass, one could use the LG as a 'local laboratory' to constrain cosmology or gravity. These constraints are unlikely to be competitive with the CMB or large scale structure (LSS), but in the case of poorly constrained or highly scale dependent MG and DE models such local systems might be useful. Observations of systems such as the LG and other galaxy clusters could eventually lead to competitive DE constraints on small scales of $\mathscr{O}(\lesssim 1\,\mathrm{Mpc})$, which are not typically probed by cosmological analyses. This is relevant primarily for scale dependent theories such as scalar-tensor theories, but not for homogeneous theories such as Λ, perfect fluids, or spatially invariant scalar fields $\phi(t)$.

Estimates for the LG mass typically derive errors from observational uncertainties on the dynamical parameters (separation, relative velocity, etc.), assuming a particular cosmology. Is the uncertainty in the argument due to the cosmology itself comparable? A consistent cosmological approach to the local group mass may be taken by first calculating the background. This will give us the age of the universe t_u from $a(t_u) = 1$ (defining the present as $t = t_u$). In addition we may obtain the full function $a(t)$, describing the entire expansion history of the universe at large. For simple models this may be calculated with the Friedmann equations for perfect fluid components. For more complicated models such as scalar-tensor theories one may require a cosmological Boltzmann code or a fixed background expansion. (We shall

used a fixed ΛCDM background expansion for scalar-tensor models and MOND.)

The timing argument is a two body problem Newtonian problem, but it can also be cast in a different light. The equations are the same for the evolution of a overdense patch of the universe with total mass $M_{\rm LG}$ evolving independently of the background. This is well known behaviour for the collapse of overdensities in structure formation. In this interpretation, the 'timing' aspect of the timing argument is that this sub-universe must have the same origin as the universe as a whole.

When non-minimally coupled scalar fields are introduced, the equations no longer hold this form. This is because the coupling to the scalar field is structure dependent – the radius and density of the objects determines the scalar field profile around an object. Typically in scalar tensor theories, the scalar field will behave in such a way as to produce little or no effect in areas of high density: this is known as 'screening'. (It is frequently achieved by having the scalar field profile fall to zero at high density, or to have the effective mass of the scalar field become very large so that its interactions are unobservably short range.) Screening is a vital component to many theories in order to modify gravity at large scales, whilst avoiding violating tests which have been made on deviations from general relativity on solar system scales [67][68]. Scalar tensor models generically produce Yukawa potentials (of the form $r^{-1}e^{-mr}$), with a theory and density dependent coupling term. In addition to this Yukawa coupling, there is a cosmological constant like effect from the energy density of the scalar field. If a scalar field is to be responsible for both dark energy and dark matter (via energy density and matter coupling respectively), then the LG mass estimates should be consistent with the baryonic mass estimates of the galaxies. If it is not, then dynamics within our local universe could become a major hurdle for theories seeking to eliminate dark matter.

5.2 Properties of the Local Group

The TA requires the relative radial separation and velocity of MW and M31, which we take to be [126]

$$r = 0.77\,\text{Mpc}, \tag{5.1}$$

$$v_r = -109.4\,\text{km}\,s^{-1}. \tag{5.2}$$

As in the previous chapter, we will deal predominantly in units of Mpc, Gyr, and $\mathrm{M}_G = 10^{12} M_\odot$. All masses in plots will be in these 'galactic' mass units.

In order to calculate M_{LG} in scalar-tensor theories, which depend on the structure of the objects and not just their mass, we need to select some additional properties for our system relating to the radius and mass of the individual galaxies. We take $M_{M31} = 2M_{MW}$ as suggested by studies such as [120], $r_{M31} = 33\,\text{kpc}$, and $r_{MW} = 30\,\text{kpc}$. In principle, these should also be varied within observational limits, but we will keep them fixed in order to focus on the scalar field theory. For dark matter free theories, the mass of the MW and of M31 is taken to be $\mathscr{O}(10^{11} M_\odot)$, with the original TA paper [92] suggesting baryonic masses of $10^{11} M_\odot$ and $4 \times 10^{11} M_\odot$ for MW and M31 respectively, and a more recent paper exploring a dark matterless MW arriving at $2.7 \times 10^{11} M_\odot$ [155]. We will therefore expect dark matter free theories to be able to produce a mass in approximately this range in order to be consistent.

5.3 The Cosmological Context of the Timing Argument with Λ

The original Timing Argument (derived in [92]) augmented with a cosmological constant (in accordance with the weak field limit of GR+Λ, see [117], [80]), hereafter referred to as TA_Λ, is highly cosmological in nature. It assumed that the MW and M31 form in close proximity in the early universe and move with the expansion of the universe, then turning around under the influence of gravity to form their present configuration. The 'timing' aspect comes from tuning the mass until

Table 5.1: A summary of the models used in this chapter, their free parameters, and the section in which they are found.

Model	Acceleration Equation	Free Parameters	Section
ΛCDM	$\ddot{r} = \ddot{r}_N + \frac{\Lambda c^2}{3} r$	Ω_Λ, H_0	5.3.1
wCDM	$\ddot{r} = \ddot{r}_N - \frac{1}{2} H_0^2 \Omega_f (1+3w) R(t)^{-3(1+w)} r$	Ω_f, w, H_0	5.4
Quintessence	$\ddot{r} = \ddot{r}_N + \frac{4\pi G}{3}[\dot{\phi}^2 + 2V(\phi)] r$	Ω_ϕ, $\dot{\phi}_0$, ϕ_*, H_0	5.5
Symmetron	$\ddot{r} = \ddot{r}_N + 4\pi \frac{C_A C_B}{M} \frac{e^{-mr}}{mr} \left(1 + \frac{1}{mr}\right)$	μ, λ, S	5.6
MOND	$\ddot{r} = -\left[\frac{\ddot{r}_N^2 + \sqrt{\ddot{r}_N^4 + 4\ddot{r}_N^2 a_0^2}}{2}\right]^{\frac{1}{2}}$	a_0, t_u	5.7

the spatial coincidence of the galaxies ($r = 0$) occurs at $t_* = t_u$. The dynamics are governed by a Newtonian equation with an additional Λ term,

$$\ddot{r} = -\frac{GM}{r^2} + \frac{\Lambda c^2}{3} r, \tag{5.3}$$

which may be obtained by taking the weak field limit of the Einstein field equations with a cosmological constant (see section 1.1.2). We can also view this equation in a cosmological context, by noting that, since the matter density ρ considered in deriving this contained only the point particles under consideration, this implicitly assumed that we are in a universe containing only the galaxy pair and dark energy. We will show that the equation of motion for two particles moving under gravity, in a dark energy filled background universe evolving according to the Friedmann equations, is identical to the form in equation 5.3.

To begin with, consider the equations in the comoving coordinate x for a particle on a background where the physical distance $r = a(t)x$ (see [56] and section 1.3.1 of this thesis),

$$\ddot{x} + 2H\dot{x} + \frac{1}{a^2}\frac{\partial \Delta\Phi}{\partial x} = 0. \tag{5.4}$$

We can find the gravitational force using the comoving Poisson equation for a spherical density distribution and a background density which is stationary w.r.t. to the

comoving coordinates:

$$x^{-2}\partial_x(x^2\partial_x\Delta\Phi) = 4\pi G(\rho-\rho_b), \tag{5.5}$$

$$\begin{aligned}\partial_x\Delta\Phi &= x^{-2}4\pi G\int(\rho-\rho_b)x^2\,dx, &(5.6)\\ &= x^{-2}4\pi G a^{-1}\int(\rho-\rho_b)r^2 dr, &(5.7)\\ &= \frac{GM}{ax^2}, &(5.8)\end{aligned}$$

which gives the equation of motion in comoving coordinates for the overdensity

$$\ddot{x}+2H\dot{x}+\frac{GM}{a^3x^2}=0. \tag{5.9}$$

Since the matter is bound up in our particles, locally the background is made up of only dark energy. From the cosmological field equations, in a universe with only a cosmological constant

$$H^2 = H_0^2\Omega_\Lambda = \frac{\Lambda c^2}{3} = \text{const.} \tag{5.10}$$

Using $\ddot{r} = a\left[H^2x+2H\dot{x}+\ddot{x}\right]$, and substituting in $\ddot{x}$ from equation 5.9, we can convert the comoving equation to our usual physical equation in r:

$$\begin{aligned}\ddot{r} &= -\frac{GM}{r^2}+H^2r,\\ &= -\frac{GM}{r^2}+\frac{\Lambda c^2}{3}r,\end{aligned} \tag{5.11}$$

which is the same as equation 5.3, even though there is no Λ term in the comoving acceleration equation. The effects of dark energy, which in ΛCDM are constant, are absorbed entirely into the scale factor $a(t)$; this is the common way to express the equations in n-body simulations and similar applications. The effect of the cosmological constant can be seen from comparisons of TA and TA_Λ with the simulations, as in the previous chapter.

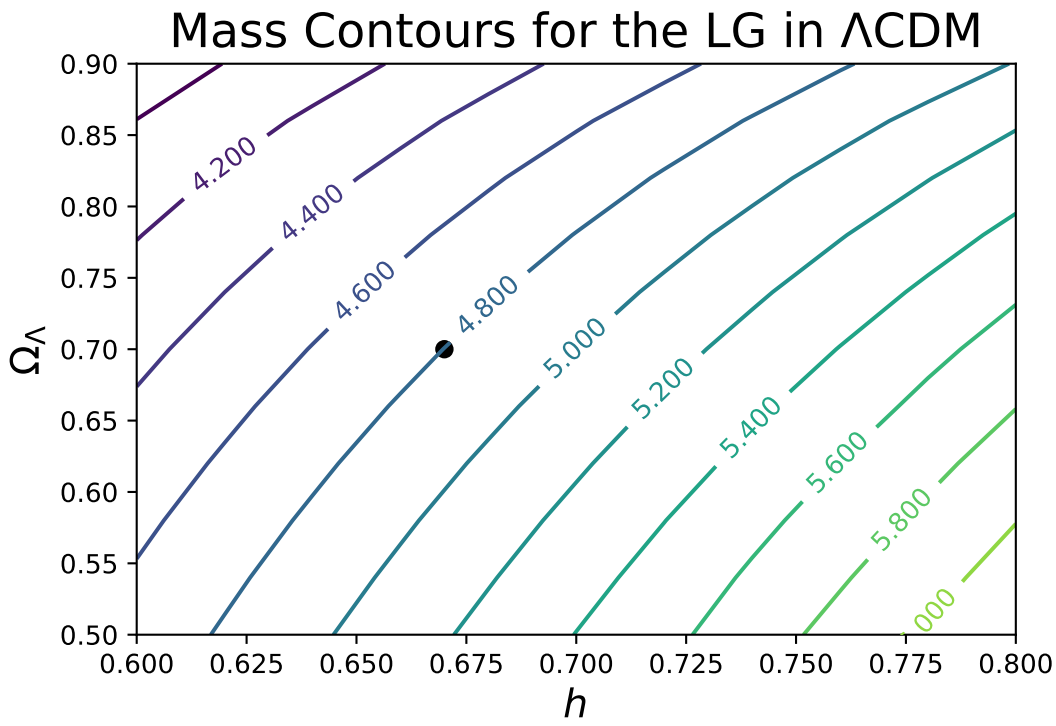
So, the equation of motion governing TA_Λ is the same as if the galaxies were evolving in a bath of dark energy: this is clearly not the case for the real universe! Indeed, the scale factor $a(t)$ here would be very different from the scale factor of a full ΛCDM universe. In fact, the matter background is highly clustered on such small scales, and most of the matter (baryonic or dark) has accreted onto bound structures in the LG. This means that the background *does* indeed have a negligible matter density and a constant dark energy density (in the ΛCDM case). This patch of universe evolves quite independently of the background expansion which governs the observable universe as a whole (it is 'decoupled' from the Hubble flow). The 'timing' aspect thus relates to the fact that the evolution of this patch of the universe must be consistent with the initial conditions of the universe as a whole i.e. the origin at $t = 0$ where the present conditions are expressed at $t = t_u$.

5.3.1 The TA in ΛCDM

In ΛCDM, the only two cosmological parameters which affect the TA_Λ are h_0 and Ω_Λ (assuming that a flat universe and negligible contributions from relativistic species, and therefore that $\Omega_\mathrm{m} = 1 - \Omega_\Lambda$). Fig 5.1 shows the LG mass estimate contours for ΛCDM varying the present values of Ω_Λ and h. Notice in particular that as Ω_Λ increases, the mass estimate actually decreases. Although we saw in the previous chapter that Λ sources an acceleration pushing the galaxies apart, which causes us to revise our mass estimate upwards, the larger effect here is that Λ increases the age of the universe, which would cause us to revise the mass estimate downwards. This demonstrates the importance of using the cosmological parameters in a consistent way in order to see the genuine effect on the dynamics. The effect of h on the mass estimate is more obvious. The age of the universe varies as $\sim H_0^{-1}$, and thus as our mass estimate should increase as h increases in order to ensure a fast enough collapse. Ω_Λ and h are also not independent, since

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H^2}, \tag{5.12}$$

Figure 5.1: Mass estimate contours for the LG using ΛCDM. The mass is given in units of $10^{12}M_{\odot}$. The dependence of the mass on the cosmological parameters comes from the impact on t_u and the Λ modification to the acceleration equation. The black dot represents the mass estimate using $\Omega_{\Lambda} = 0.7$ and $h_0 = 0.67$, which is a typical mass estimate for a Planck-like cosmology.



and thus it is $H^2\Omega_{\Lambda}$ which enters into the acceleration equation. Nevertheless, Ω_{Λ} is much more useful for cosmology than Λ itself, and it is more useful to look at the impact of the parameters which are widely used. We can see from Fig 5.1 that varying h over the range of different estimates made by various cosmological probes could yield significant changes to the LG mass estimate. Taking h in the range $[0.632, 0.764]$ (to take the extremes of [75], the lowest value being the lower bound for a study of Megamasers [79], and the higher value being the upper bound for a Cepheid study [78]), and a fixed $\Omega_{\Lambda} = 0.7$, we find the mass falls in the range $[4.55 \times 10^{12}M_{\odot}, 5.43 \times 10^{12}M_{\odot}]$. This level of uncertainty is comparable to the uncertainties in the observable dynamics or the scatter in the regression models in tables 4.4 and 4.5.

5.4 Timing Argument for Perfect Fluids

To cast the TA for different theories of gravity, we need to look at the weak field limit of general relativity including the relevant dark energy model. Perfect fluid models have one extra free parameter, which is their equation of state $w = \frac{P_f}{\rho_f}$, where $w_\Lambda = -1$. The strength of DE (determined by the density ρ_f) and the equation of state (which may be a function of time) will both affect the acceleration equation. The equation of state enters into both the energy-momentum tensor (which therefore affects the gravitational strength at a given time) and also the variation in density of dark energy (which is constant for Λ).

The weak field quasi-static (non relativistic) limit of general relativity gives us

$$\nabla^2\Phi = 4\pi G(\rho_m + (1+3w)\rho_{\rm f}). \tag{5.13}$$

Since ∇^2 is a linear operator, we can split the potential into two parts Φ_m and $\Phi_{\rm DE}$, which correspond to the matter and dark energy potentials; these then lead to separate contributions to the acceleration equation $\ddot{r} = \ddot{r}_m + \ddot{r}_{\rm DE}$. The matter potential leads to the familiar Newtonian expression for a spherical density distribution. The contribution to the acceleration equation can therefore be determined by integrating only the fluid part of the equation:

$$\nabla^2\Phi_{\rm DE} = 4\pi G(1+3w)\rho_f. \tag{5.14}$$

We will assume that w is constant, and that ρ_f is not spatially varying (i.e. dark energy is non-clustering). The contribution to the acceleration is therefore:

$$\ddot{r}_{\rm DE} = -\nabla\Phi_{\rm DE} = -\frac{4\pi G(1+3w)}{r^2}\int_0^r \rho_f(r')r'^2\,dr', \tag{5.15}$$

which depends on the equation of state and the integrated energy density of the dark energy fluid. The case of a cosmological constant is easily recovered when $w = -1$

and $\rho_f = \rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$. Integrating, we have:

$$\ddot{r}_{\rm DE} = -\frac{4\pi G}{3}(1+3w)\rho_f r. \tag{5.16}$$

Since our dark energy is not clustered, the density is determined entirely by the background of the universe at large,

$$\rho_f = \rho_{f,0}\, a^{-3(1+w)}. \tag{5.17}$$

The acceleration equation is therefore:

$$\ddot{r} = -\frac{GM}{r^2} - \frac{1}{2}H_0^2\Omega_f(1+3w)a(t)^{-3(1+w)}r. \tag{5.18}$$

In order to consistently use dark energy models in the TA, we must also know how it affects the age of the universe and, in the case of a perfect fluid, its expansion history $a(t)$. Both can be readily calculated from the Friedmann equations using cosmological parameters H_0 and $\Omega_{\rm DE,0}$ (assuming $\Omega_{\rm m,0} = 1 - \Omega_{\rm DE,0}$) to set the conditions at the present time, and then integrating back until $a(t_0 - t_u) = 0$,

$$H^2 = H_0^2 \sum_i \left[\Omega_{i,0}\, a^{-3(1+w_i)}\right]. \tag{5.19}$$

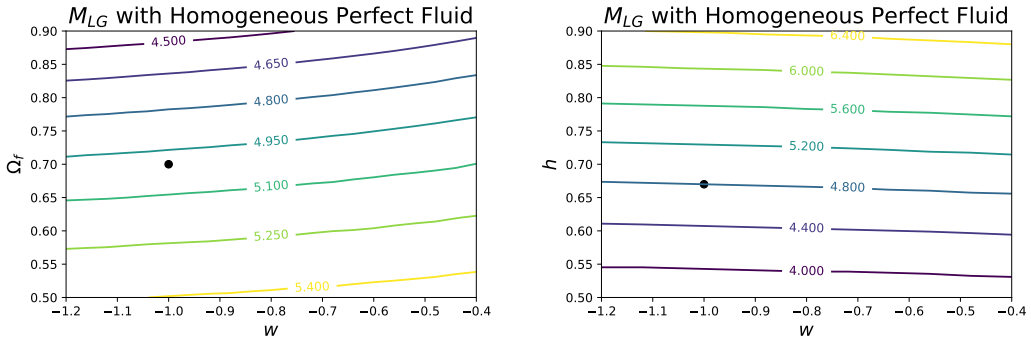
There is an analytic expression for a(t) for the case of a two component (matter and dark energy) model with constant w:

$$H_0 t = \frac{2\ln\left(\sqrt{\Omega_f(\Omega_m + \Omega_f a^3)} + \Omega_f a^{\frac{3}{2}}\right)}{3\sqrt{\Omega_f}} - \frac{2\ln(\sqrt{\Omega_m\Omega_f})}{3\sqrt{\Omega_f}}. \tag{5.20}$$

The age of the universe is the difference in time between $a = 1$ and $a = 0$. (Alternatively, it can be readily obtained by numerical integration.) The mass is then numerically calculated from (r, v_r, t_u), using this background cosmology to inform the acceleration equation 5.18.

Figs 5.2 shows the LG mass estimate as a function of Ω_f, w, and h for a spatially homogeneous perfect fluid with constant equation of state. The LG mass is

Figure 5.2: Mass estimate contours for the LG using a spatially homogeneous perfect fluids with constant w. The black dot corresponds to ΛCDM with $\Omega_\Lambda = 0.7$ ($\Omega_f = 0.7$, $w = -1$) and $h = 0.67$.



not very sensitive to the equation of state at all, even when deviations from ΛCDM are extreme. Like the case with Λ, there are two opposing effects in play here. Decreasing w decreases the age of the universe, but also diminishes the dark energy expansion effect in the acceleration equation. These effects are largely canceling each other out in the mass calculation.

5.5 Minimally Coupled Scalar Field

A scalar field which does not directly couple to gravity or to matter can produce a dark energy effect. If it has only time dependence then it can be easily inserted into the gravitational equations using the energy-momentum tensor of the scalar field. The scalar field evolves according to a potential function $V(\phi)$; this must be tuned to produce the effect that we want. Minimally coupled scalar fields are typically invoked in inflation theories, although they also form the basis of quintessence theories of dark energy. Quintessence was introduced in [74], and a review can be found here [73]. In our case we will make use of straightforward properties of the scalar field which may be found in textbooks such as [26]. Attempts to constraint the quintessence potential have been made in [72], although the uncertainties remain large, and higher than linear or quadratic terms in the potential remain unconstrained.

Since we are assuming a spatially homogeneous scalar field across the entire universe, the scalar field value (like the density of the perfect fluid in the previous

section) must be calculated taking the evolution of the universe into account. Using the Klein-Gordon equation (see section 1.2.1 for details)

$$\Box\phi = g^{\mu\nu}\nabla_\mu\nabla_\nu\phi = \partial_\phi V(\phi), \tag{5.21}$$

and using the FLRW metric and the fact that ϕ is a function of t only, one obtains:

$$\ddot{\phi} + 3H\dot{\phi} - \partial_\phi V(\phi) = 0. \tag{5.22}$$

In order to calculate the evolution of the scalar field then, we need to have $V(\phi)$, ϕ_0, and $\dot{\phi}_0$.

The acceleration equation for a universe with a scalar field can be found using the expression for the scalar field energy momentum tensor, which is itself obtained by varying the action in equation 1.25 with respect to the metric:

$$T^{\phi}_{\mu\nu} = \partial_\mu\phi\,\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}\partial_\sigma\phi\,\partial^\sigma\phi - V(\phi)\right). \tag{5.23}$$

Inserting this into the Einstein field equations gives the acceleration equation. Using the Newtonian limit (setting $c = 1$)

$$\nabla^2\Phi = 4\pi G\left[\rho - 2\dot{\phi}^2 - 2V(\phi)\right]. \tag{5.24}$$

It is easy to show that this is equivalent to inserting $\rho_\phi - 3p_\phi$ from equation 1.29 into equation 5.13 i.e. that the field in this limit acts as a perfect fluid with variable w_ϕ (given by equation 1.31). The limiting case of a cosmological constant is found when $\dot{\phi} = 0$ and therefore:

$$8\pi G V(\phi) = \Lambda. \tag{5.25}$$

Integrating equation 5.24 and bearing in mind that the scalar field has no spatial dependence, one obtains

$$\ddot{r} = -\frac{GM}{r^2} + \frac{8\pi G}{3}\left[\dot{\phi}^2 + V(\phi)\right] r. \tag{5.26}$$

We use an exponential form for the quintessence potential, choosing:

$$V(\phi) = V_0 e^{-\frac{\phi}{\phi_*}}. \tag{5.27}$$

We may set $\phi_0 = 0$ without loss of generality, as the effect of starting ϕ at different points can be absorbed into V_0. ϕ_* sets the gradient of the potential, and $\dot{\phi}$ sets its initial trajectory. We calculate V_0 by relating it to Ω_ϕ:

$$\Omega_\phi = \frac{8\pi G\rho_\phi}{3H^2} \implies V_0 = \frac{3H^2\Omega_\phi}{8\pi G} - \dot{\phi}_0^2. \tag{5.28}$$

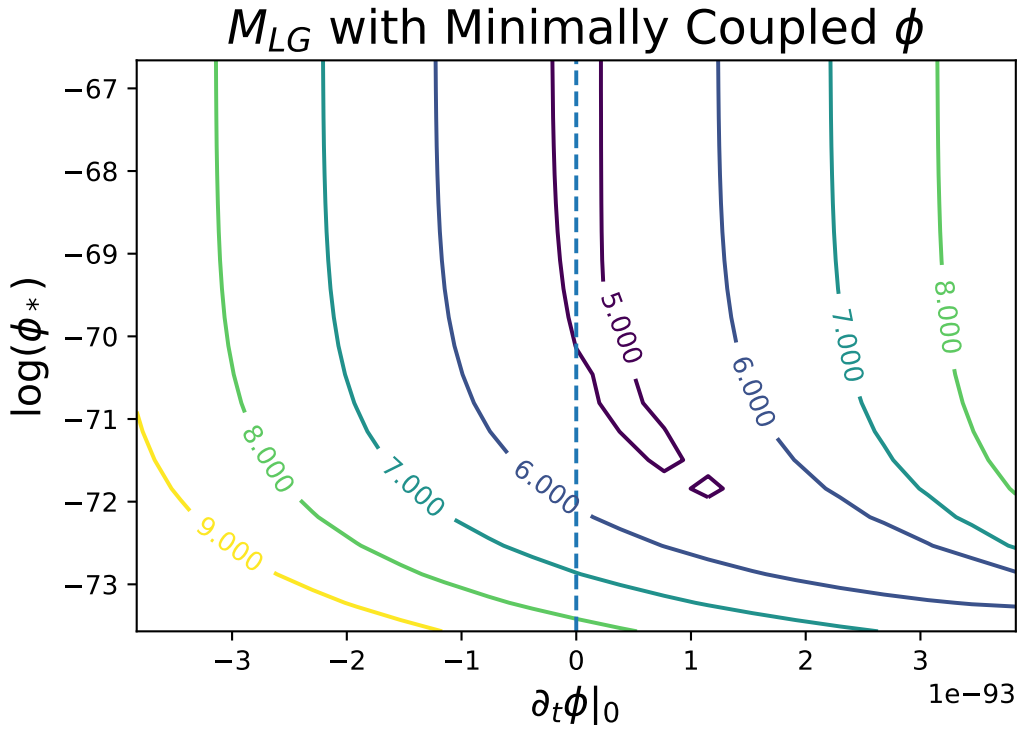
With these initial conditions, the model is fully governed by Ω_ϕ, ϕ_*, and $\dot{\phi}_0$. Fig 5.3 shows the variation in the mass estimate with $\dot{\phi}_0$ and ϕ_*. We can see that the LG mass is insensitive to ϕ_* when the parameter is large enough. This is because the potential is flattened, and its contribution becomes roughly constant for all $\phi_* \gg \phi$. In this case the change to the mass is also almost symmetric in $\dot{\phi}$. The field velocity only enters into the acceleration equation as $\dot{\phi}^2$, so if the potential is extremely flat and thus making little impact on the field velocity then positive and negative velocities will be almost indistinguishable. When $\phi*$ is small enough, then the potential is rapidly varying and it becomes the dominant part of the dynamics, with the initial scalar field velocity $\dot{\phi}_0$ quickly becoming overwhelmed by the roll down the potential.

5.6 Coupled Scalar Fields

The scalar field need not act independently of the matter. Coupled scalar fields – scalar-tensor theories – interact with the matter or the spacetime curvature directly to produce spatially inhomogeneous fields. A thorough introduction to scalar-tensor theories can be found in [29], and a broad but less detailed review in [67].

The Newtonian limit of such a scalar field theory is determined by the matter coupling $\Omega(\varphi)$ and the evolution of the scalar field. For a given theory the scalar field is determined by a Klein-Gordon equation with an effective potential which, in the Einstein frame, depends on the matter density. The masses and radii of MW

Figure 5.3: Mass estimate contours for the LG with a spatially homogeneous scalar field with an exponential potential $V(\phi) = V_0 e^{-\frac{\phi}{\phi_*}}$. The dashed line represents $\dot{\phi} = 0.0$, which approached ΛCDM in the limit $\phi_* \to \infty$. Scalar field quantities are presented in Planck units.



and M31 are likely different, and therefore the profile of the scalar field around each of them may be different. Therefore we cannot combine their mass in the acceleration equations quite so simplistically, and the equation will involve both masses individually.

Scalar fields may be used to model either dark energy or dark matter (as in the symmetron field in [65]) which means that some scalar-tensor theories may be comparable to MOND. The energy density of the scalar field can act in the same way as a cosmological constant and produce a universal repulsion; this is true even for fields which produce an increased gravitational potential which can mimic dark matter. It would certainly be interesting if the dark aspects of cosmology could both be illuminated by the additional of a single scalar field! In general however, the potential of a scalar field theory must often be augmented with an additional constant to mimic Λ.

5.6.1 The Klein Gordon Equation for Scalar Fields

Perhaps the first thing we need to understand is the field equation for the scalar itself. This is often referred to as the 'Klein-Gordon equation' of the scalar field, by analogy with the famous equation for free fields. The field will be subject to some effective potential which, in the case of coupled scalar fields, will be a function of the scalar field and either the curvature (Jordan frame) or the matter density (Einstein frame). Let us assume we have a scalar field ϕ, which is subject to a scalar field equation:

$$\Box^2\phi + \partial_\phi V_{\rm eff}(\phi) = 0. \tag{5.29}$$

Provided that the effective potential $V_{\rm eff}$ has a minimum at some ϕ_0, then we can say

$$\partial_\phi V_{\rm eff}(\phi_0) = 0. \tag{5.30}$$

We can expand the scalar field around the minimum such that $\phi = \phi_0 + \varphi$. The linearised equation then reads:

$$\Box^2\varphi + \partial^2_\phi V_{\rm eff}(\phi_0)\varphi = 0. \tag{5.31}$$

We can then identify $\partial^2_\phi V_{\rm eff}(\phi_0)$ with an effective mass $m^2_{\phi_0}$, and rewrite the equation as

$$\left(\Box^2 + m^2_{\phi_0}\right)\varphi = 0, \tag{5.32}$$

which makes the connection to the Klein-Gordon equation. We can already see a key feature of coupled theories: the scalar field mass becomes density (or curvature) dependent. This is what allows the scalar field to 'hide' in high density environments, evading laboratory or solar system detection.

5.6.2 Generic Spherically Symmetric Solutions for Uniform Spheres

Spherically symmetric solutions for the scalar field equation have been discussed in [67][68][64]. Equation 5.32 for the scalar field perturbation will in general have

complex wave like solutions, although in the quasi-static limit there are growing and decaying modes. We shall find solutions of the form

$$\varphi = Cx^a e^{bx}, \tag{5.33}$$

for real C, a, and complex b (since φ is a real scalar field). In the quasi-static case where we ignore time derivatives, $\Box^2 \to -\nabla^2$, and we may expand equation 5.32 for a spherically symmetric case as:

$$\varphi'' + \frac{2}{r}\varphi' - m_{\phi_0}^2 \varphi = 0, \tag{5.34}$$

where prime indicates derivatives with respect to r. If we change coordinates to $x = m_{\phi_0} r$, then the equation can be written:

$$\varphi'' + \frac{2}{x}\varphi' - \varphi = 0, \tag{5.35}$$

where prime now indicates derivatives with respect to the dimensionless scaled coordinate x. Using our trial solution:

$$\varphi = Cx^p e^{qx}, \tag{5.36}$$

$$\varphi' = \left[\frac{p}{x} + q\right]\varphi, \tag{5.37}$$

$$\varphi'' = \left[\frac{p^2 - q}{x^2} + \frac{2pq}{x} + q^2\right]\varphi. \tag{5.38}$$

Inserting these expressions into equation 5.35 and collecting terms in factors of x yields the consistency equations:

$$q^2 - 1 = 0, \tag{5.39}$$

$$p(p+1) = 0, \tag{5.40}$$

$$2pq + 2q = 0, \tag{5.41}$$

which can be solved by $p = -1$ and $q = \pm 1$. We are then generically led to solutions of the form:

$$\phi = \phi_0 + \varphi = \phi_0 + \frac{1}{m_{\phi_0} r}\left[Ae^{-m_{\phi_0} r} + Be^{m_{\phi_0} r}\right]. \tag{5.42}$$

This solution can be applied around the minimum of a generic potential form. It is the boundary conditions which lead us to the different internal and external solutions. At the centre we must have, by spherical symmetry, $\varphi' = 0$. We also insist that the scalar field should tend to the free space minimum of the potential ϕ_{bg} in the limit of large r, and thus our functional forms are determined by:

$$\varphi'(r = 0) = 0, \tag{5.43}$$

$$\lim_{r\to\infty} \varphi(r) = \lim_{r\to\infty} \varphi'(r) = 0. \tag{5.44}$$

This leads to solutions of the form

$$\phi = \begin{cases} \phi_c + D\frac{\sinh(m_c r)}{m_c r} & \textit{internal}, \\ \phi_{bg} - C\frac{\exp(-m_{bg} r)}{m_{bg} r} & \textit{external}. \end{cases} \tag{5.45}$$

Where ϕ_c is the scalar field value which minimises the internal potential $\rho \neq 0$, and ϕ_{bg} is the field value which minimises the potential outside. The crucial constant C – which determines the amplitude of the scalar field perturbation far from the object and therefore also the modification to gravitational attraction – may be determined by matching the internal solution to the external solution at the boundary of the object. Fixing both ϕ and ϕ' fixes both D and C (although D holds no real significance on its own). The obtained coupling strength, C, will depend on $M_{\rm LG}$ since the internal solution depends on the structure of the object including its density and radius.

Considering a uniform sphere with radius R, we can join the solutions at the boundary of the sphere. This leads to an amplitude of

$$C = (\phi_{bg} - \phi_c)e^{m_{bg}R}\left[\left(1 + \frac{1}{m_{bg}R}\right)\frac{1}{m_c R\coth(m_c R) - 1} + \frac{1}{m_{bg}R}\right]^{-1}. \tag{5.46}$$

In the case where $m_{bg}R, m_cR \ll 1$ we may Taylor expand the above expression to arrive at the simpler (and more intuitive) expression:

$$C \approx \frac{\Delta\phi}{3} x_{bg} x_c^2, \tag{5.47}$$

where $x_{bg} = m_{bg}R$, $x_c = m_cR$, and $\Delta\phi = \phi_{bg} - \phi_c$. Although this is only valid in the perturbative regime for a highly idealised object, it can help us understand how the coupling to objects depends on the internal structure through changes to the minimum ground state and effective mass of the scalar field inside the object, which in turn depends on its density and radius.

5.6.3 Solutions for Strongly Perturbed Scalar Fields in Uniform and Non-Uniform Spheres

Should we wish to relax our assumptions and calculate solutions for more complex object or a more strongly perturbing regime, we must resort to numerical analyses in the general case. Numerical methods are used in [65] to calculate the field inside a non-spherical galaxy, and [64] considers strongly perturbing solutions. In order to numerically integrate the solution, one must then set the initial values of the scalar field at the centre of the object to be:

$$\phi(r=0) = \phi_*, \tag{5.48}$$

$$\phi'(r=0) = 0. \tag{5.49}$$

Different values of ϕ_* will lead to radically different behaviours, and we seek the numerical solution which obeys the asymptotic conditions outlined above in the analytic solution. We know that at large r we must have the exponentially decaying solution if we have a constant background density (which we may take to be zero) since we will eventually be close enough to ϕ_{bg} for the perturbative solution to be valid. Clearly the combination $\frac{\phi_{bg}-\phi}{\phi'}$ is independent of the constant amplitude C (which is what we want to find). So at a radius large enough that we are confident that the asymptotic solution applies, we may find the correct numerical solution by

tuning ϕ_* such that $\frac{\phi_{bg}-\phi}{\phi'}$ gives the correct value. This then uniquely determines the amplitude and thus gives the far field solution even for complicated and highly structured objects. The main barrier to this approach is the numerical stability of the solutions, which may be difficult to model for strongly perturbed fields. We shall not consider strongly perturbing fields further for this reason.

5.6.4 Fifth Force Interaction Between Two Objects

We have seen in the introduction (equation 1.101) that a test particle under scalar-tensor gravity will follow a modification to the geodesic in the Einstein frame (or, equivalently, geodesics in the Jordan frame) based solely on the form of the scalar field and the matter coupling at that point. The additional acceleration due to the scalar field is:

$$\ddot{x}_\phi = -\nabla \ln(\Omega(\phi)). \tag{5.50}$$

For an extended object however, the scalar field itself is affected by the matter density in a structure dependent way, and so for a realistic object we must also know the internal solution of the scalar field for the object (or objects) before we can understand how the object will move.

The force on an object in general is the rate of change of momentum of that object $\vec{F} = \dot{\vec{p}}$. The derivation of this fifth force in the case of a Chameleon field around a small uniform sphere was given in Appendix B of [64]. We shall therefore only highlight the key aspects, and note where our case deviates from theirs. In fact, the derivation is largely applicable with only a small modification to the force term. Two spherical objects at large separation, A and B, are considered and the force on B due to A is calculated.

The scalar field equation is, in general, non-linear ($\mathbf{O}\phi = \Box^2\phi + \partial_\phi V(\phi)$ is a non-linear operator). This means that many body solutions are not generally superpositions of single body solutions. It is possible to superpose solutions in the case where the linearisation of the field equations is valid i.e. when the field is written $\phi = \phi_0 + \varphi$. In this case $\varphi_{A,B}$ individually satisfy the linear equations and thus

$$\phi = \phi_0 + \varphi_A + \varphi_B. \tag{5.51}$$

is a solution to the field equations. (Note that the field *perturbations* are added, not the entire field solutions.) In this case, the form of the scalar field is simple to calculate, and the contributions from each object are well separated.

Inside the body we will assume that the scalar field is dominated by the body itself (as it is sensitive to the local density) so that the coupling term C_B (which depends on the interior behaviour of the field in object B) is not sensitive to body A and vice versa. This allows us to use the scalar field solutions in equation 5.45 for the field perturbations φ_A and φ_B, centred around their respective objects.

Expressing the total momentum of object B as an integral over a spherical volume around object B and differentiation, the force may be expressed as a surface integral around the spherical object,

$$F_i = \dot{P}_i = -\int_V \partial_j \tau^i_j d^3x = -\int_S \tau^j_i n_j dS, \tag{5.52}$$

where i, j run over spatial indices, and $\tau_{\mu\nu}$ is the energy momentum tensor for matter, the scalar field, and gravity (second order parts of the metric fluctuations). The gravitational part of the theory is unchanged in the Einstein frame and thus returns our usual Newtonian force law. The surface is drawn just outside of a bound object, so the matter terms are negligible here. The fifth force must come from the energy-momentum of the scalar field, which is the only part of $\tau_{\mu\nu}$ which does not appear in the standard GR case. To calculate the scalar field contribution to $\vec{F}$, we recall the energy momentum of the scalar field is

$$T^{(\phi)}_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}(\nabla_\sigma\phi)(\nabla^\sigma\phi) - V(\phi)\right], \tag{5.53}$$

and the gradient of our scalar field solution:

$$\nabla_i\varphi = \begin{cases} \frac{Ce^{-m_{bg}r}}{m_{bg}r}\left[r^{-1}+m\right]\frac{x_i}{r} & \\ \frac{C}{m_{bg}}\frac{x_i}{r^3} & r \ll m_{bg}^{-1}. \end{cases} \tag{5.54}$$

Since the surface is drawn around object B, and the coordinates are centred on B, we

may use the latter approximation for $\nabla_i \varphi_B$ provided the radius of the object is much less than the characteristic scale of the force (which should be the case). The full derivative will need to be used later for the gradient of the scalar field originating from A as the distance from A to B is much larger than the radii of the objects themselves. Due to the geometry of the setup most of the contributions cancel out when integrated over the sphere, and only one term remains after the integration. We have:

$$F_i = -4\pi \frac{C_B}{m_{bg}} \partial_i \varphi_A. \tag{5.55}$$

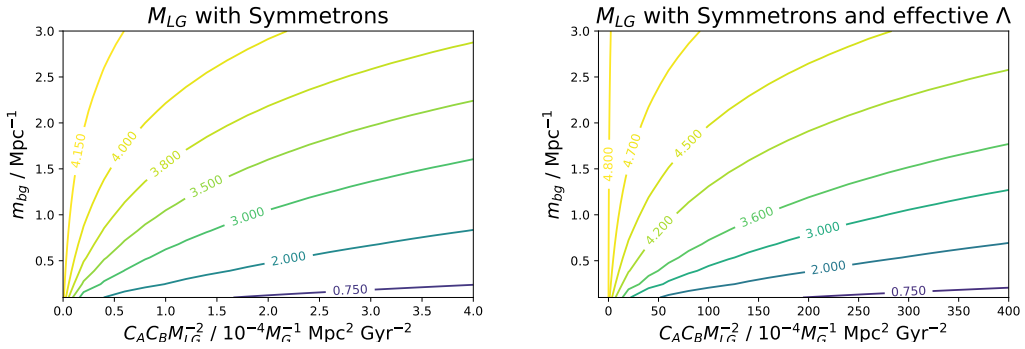
Now switching our coordinate system and letting r be the radial separation between objects A and B, and inserting the gradient of the field φ_A from equation 5.54, we find the radial force law:

$$F_\phi = -4\pi C_A C_B \frac{e^{-m_{bg} r}}{m_{bg} r} \left(1 + \frac{1}{m_{bg} r}\right). \tag{5.56}$$

The functional form of this is largely as one would expect from heuristic considerations (see [68]), and involves a Yukawa force with a coupling to each body depending on the degree of screening which is exhibited by each. This attractive force can act as the source of scale dependent modified growth.

Although [64] derives a force law which is r^{-2}, and thus at large distances is indistinguishable from an additional mass, we are forced to confront the exponential decay of our fifth force. In the case of a dark matter free theory, this can lead to a discrepancy in the effective mass at short and large distances. MOND theories, which seek to stabilise rotation curves, are known to lead to low mass estimates inconsistent with the baryonic mass estimates of the galaxies when applied in the TA [66]. The exponential suppression in the far field force could help to remedy this discrepancy. On the other hand, if the exponential suppression is too large, then one may find that the fifth force interaction which is making up for the lack of dark matter inside the galaxy falls too quickly at large distances to be consistent with Mpc scale dynamics without additional mass. Given the success of dark matter across a wide range of different observables, any fifth forces seeking to replace dark matter

Figure 5.4: Mass estimate contours for the LG using a generic spherically symmetric scalar field solution and varying the coupling and background effective mass of the scalar field.



without conforming to a r^{-2} power law may find it difficult to match observations on all scales.

5.6.5 M_{LG} in coupled theories and assumed properties of the Local Group

Without selecting a particular theory, we may already use this generic form to investigate the effect on the LG, as is shown in Fig 5.6. This will apply to any scalar field theory which generates these couplings and background effective masses. The ΛCDM limit is as $C_A C_B \to 0$ or $m_{bg} \to \infty$. The previous derivation of the fifth force does not include the acceleration produced by the energy density of the scalar field between the two objects (analogous to the scalar field, perfect fluid, and Λ in the previous sections). In general this potential must be tuned to produce a Λ like effect, and so we also explore the acceleration equation using a coupled scalar field with a cosmological constant Λ.

5.6.6 The Symmetron Model

Specific models may be explored by calculating the relevant scalar field effective mass $m_{\phi_{bg}}$, the background scalar field ϕ_{bg} and using the appropriate coupling for the theory. We consider a symmetron model here, based on the parameters in [65]. In this case, the symmetron was introduced to stabilise rotation curves without the need for dark matter. Although this was based on numerical solution for observed galaxy profiles in cylindrical models, for the sake of being about to calculate a force

law we will restrict ourselves for now to uniform spheres.

The symmetron model has a quartic potential, of the form:

$$V(\phi) = V_0 - \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4, \tag{5.57}$$

where $\mu^2, \lambda > 0$. The effective potential is coupled to the matter distribution, although far from the object the matter distribution is negligible. The minimum of the potential is at

$$\phi_{bg} = \pm\frac{\mu}{\sqrt{\lambda}}. \tag{5.58}$$

The effective mass at the minimum, $m_{\phi_{bg}}$, is given by:

$$m_{\phi_{bg}} = V''(\phi_{bg}) = 2\mu^2. \tag{5.59}$$

The final ingredient we require is the matter coupling, which is given by

$$A(\phi) \approx 1 + \left(\frac{\phi}{S}\right)^2 + \mathcal{O}\left[\left(\frac{\phi}{S}\right)^3\right]. \tag{5.60}$$

As an aside, it is useful to look at the potential of such a scalar field background:

$$V(\phi_{bg}) = V_0 - \frac{\mu^2}{4\lambda}. \tag{5.61}$$

If the gradient of the scalar field is negligible compared to the potential (likely to be true over areas where the density is roughly uniform) then this means the energy density of the scalar field $T^{\phi}_{\mu\nu}$ is dominated by the potential, which is always $< V_0$ by definition. This means that in order to produce a dark energy effect, symmetrons would be required to have a positive definite V_0, which is, mathematically at least, equivalent to adding a cosmological constant into the theory. We will thus also consider the behaviour of a symmetron model with Λ.

The more general case in the presence of a matter density ρ_m is required to understand the contribution of the scalar field to gravitation. The crucial properties are the ground state of the scalar field ϕ_g and the effective mass there m_g.

$$V(\phi,\rho) = V_0 + \frac{1}{2}(\mu_\rho^2 - \mu^2)\phi^2 + \frac{1}{4}\lambda\phi^4, \tag{5.62}$$

where $\mu_\rho^2 = \frac{\rho}{S^2}$. Note that when $\mu_\rho > \mu$ (high density environments) then the potential has only one stationary point, a minimum at $\phi = 0$. At lower density, the central stationary point becomes a local maximum, and there are two minima at

$$\phi_0 = \pm\sqrt{\frac{\mu^2 - \mu_\rho^2}{\lambda}}, \tag{5.63}$$

which tends to the background solution (Eqn. 5.58) for $\rho_m = 0$. The mass is given by:

$$m^2 = 2(\mu^2 - \mu_\rho^2). \tag{5.64}$$

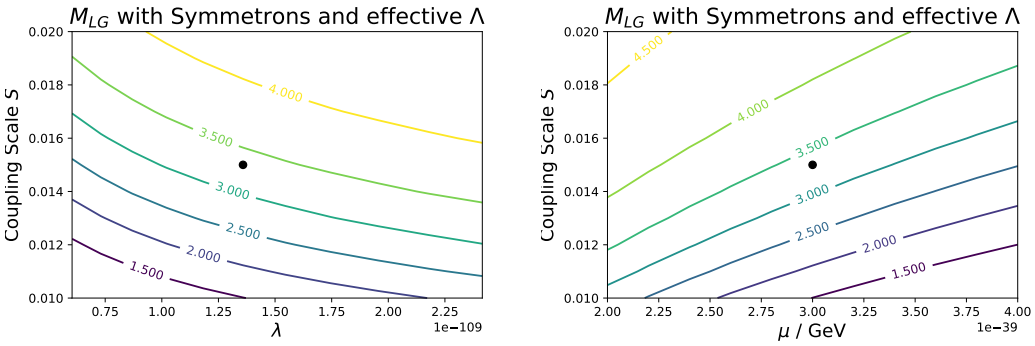
In the case where $\phi_0 = 0$ ($\mu_\rho > \mu$), then the mass is $(\mu_\rho^2 - \mu^2)^{\frac{1}{2}}$.

Using the expression for the coupling (Eqn. 5.46), we can see that as the density of the object increases, the coupling will decrease. (Although $\Delta\phi$ will increase, the $[x\coth(x) - 1]^{-1}$ term grows more quickly.) This leads to very dense objects decoupling from symmetron fifth-force effects.

The free variables here are then μ, λ, and S; we may take Ω_Λ and the background expansion to be typical of ΛCDM. One may also wish to vary the structural properties of the galaxies which are not necessarily well observed, such as the effective radius of the halo, but we shall not do this here.

Screening is an important part of modified gravity models. Contrary to many models, the effective mass of the scalar field is actually lower in high density environments, so that the force becomes very long ranged. Screening in the symmetron model is instead achieved by having the scalar field value drop to zero in high density environments. This reduces the conformal coupling $A(\phi) \to 1$, which returns GR. (There may remain a vestigial potential $V(\phi = 0) = V_0$ which would act as a cosmological constant.) When the local density is low enough the minimum of the potential rises so that $|\phi_0| > 0$, and the effects of the field may become apparent. In order for the symmetron to have a significant effect on the motions of the

Figure 5.5: Mass estimate contours for the LG using a symmetron model and varying the free parameters in the potential function. The MW and M31 have been taken to be uniform spheres with radii 30 and 33 kpc respectively, with $M_{M31} = 2M_{MW}$. The black dot corresponds the the fiducial values in [65]. The mass estimate is significantly lower than in GR due to the attractive scalar field fifth force; nevertheless it is still nearly an order of magnitude too large to be accounted for by the baryonic mass of the two galaxies.



galaxies, we require the field to be poorly screened on the scale of the galaxy. It can nevertheless be well screened in even denser environments, such as the solar system.

Using the symmetron model we can explore the effects of the symmetron potential parameters directly, as in Fig 5.5. Notice that the estimated mass for the values given in [65] (shown as the black dots on Fig 5.5) is much larger than estimates of the mass of MW and M31 in standard matter, which could indicate a significant inconsistency in using this theory of gravity to eliminate dark matter. It should be noted that more detailed modelling would need to be used to confirm this, such as modelling the galaxies as disks with realistic density profiles, and abandoning spherical symmetry in favour of cylindrical. (This leads to significant complications if the disks are not co-planar.) Despite the shortcomings of our model, this places significant pressure on this symmetron model as an alternative to dark matter.

5.7 Modified Newtonian Dynamics (MOND)

MOND is another alternative to dark matter, which was first proposed in 1983 [69]. Since its inception a large number of variants have been produced which produce the same limiting case dynamics in different ways. Its best known achievements are fitting galactic rotation curves and providing an explanation for the Tully-Fisher

relation. Nevertheless, a gravitational law which stabilises a galactic rotation curve does not necessarily have desirable long range behaviour; the non-dark matter of the MW and M31 must be able to generate a similar far field acceleration to dark matter in order to produce the observed dynamics. One might immediately suspect that this could be an issue, since in the deep-MOND regime (valid where the galaxies are well separated and the gravitational interaction is very weak – much weaker indeed than the gravity at the edge of the disk) one has $a \approx \sqrt{a_N a_0}$ i.e. the geometric mean of the Newtonian acceleration and the fixed acceleration scale a_0. This then naturally produces $a \sim \sqrt{r^{-2} a_0} \propto r^{-1}$. We therefore expect the long range gravitational acceleration to be excessive, and thus the $\mathrm{TA}_{\mathrm{MOND}}$ mass estimate to be too small. This is consistent with [66] for example, who find that MOND implies a close encounter in the past between MW and M31. It would, however, need to remain sufficiently separated to avoid tidal disruptions to either galaxy, as these are not observed.

The results in MOND will depend upon the universal acceleration scale, which can be independently determined from galaxy rotation curves. Calculating a consistent cosmological history in a non-GR theory such as MOND is non-trivial. We are not in a position to calculate the age of the universe in a fully relativistic formulation of MOND, although we can test the MOND model using a sensible range of ages for the universe $11 - 15$ GYr.

5.7.1 Acceleration Equations in MOND

The non-relativistic MOND theory has a non-linear modified Poisson equation [70]:

$$\nabla \cdot \left[\mu \left(\frac{a}{a_0} \right) \nabla \Phi \right] = 4\pi G \rho = \nabla \cdot \nabla \Phi_N. \tag{5.65}$$

From this we may infer:

$$\mu \left(\frac{a}{a_0} \right) \nabla \Phi = \nabla \Phi_N + \nabla \times F, \tag{5.66}$$

where the curl field $\nabla \times F$ is taken to be zero to match the boundary conditions that $\nabla \times F = 0$ at $r \to \infty$ and immediately around a spherical object (since the ac-

celeration field should be radial), as well as keeping the correct limiting behaviour for the Newtonian and deep-MOND regimes. The 'standard' MOND term uses the parameterisation

$$\mu\left(\frac{a}{a_0}\right) = \left[1+\left(\frac{a_0}{a}\right)^2\right]^{-\frac{1}{2}}. \tag{5.67}$$

Rewriting $-\frac{GM}{r^2} = a_N$ we arrive at the acceleration equation

$$a^4 = a_N^2 a^2 + a_N^2 a_0^2, \tag{5.68}$$

which is just a simple quadratic equation. Given that the determinant $a_N^4 + 4a_N^2 a_0^2 > a_N^4$ there is one positive solution for a^2. We finally have

$$a = -\left[\frac{a_N^2 + \sqrt{a_N^4 + 4a_N^2 a_0^2}}{2}\right]^{\frac{1}{2}}, \tag{5.69}$$

for the standard interpolating function.

There is also a 'simple' interpolating function, which leads to a slightly different solution. We will investigate the effects of both. The simple interpolating function is

$$\mu\left(\frac{a}{a_0}\right) = \left[1+\frac{a_0}{|a|}\right]^{-1}, \tag{5.70}$$

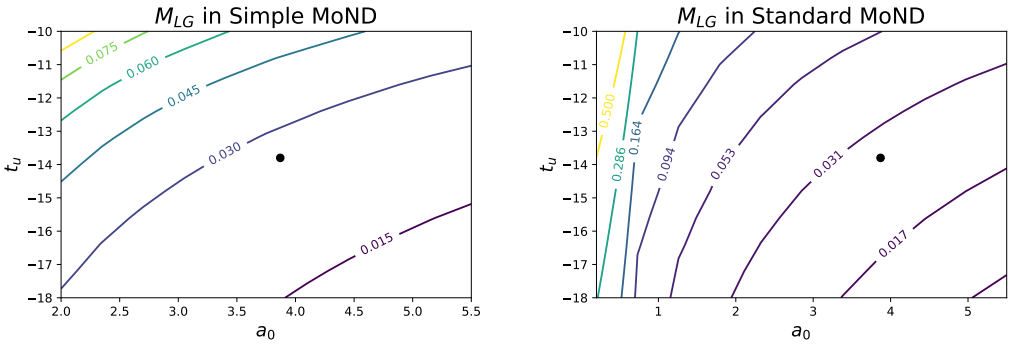
which leads to a closely related acceleration equation:

$$a = -\left[\frac{|a_N| + \sqrt{|a_N|^2 + 4|a_N||a_0|}}{2}\right]. \tag{5.71}$$

We may investigate the mass of the LG when varying a_0 and t_u. (We do not vary the cosmological parameters directly in this case as it is not clear what the cosmological implications of non-relativistic MOND is.)

In the case of a LG analysis, both interpolating functions give similar results as we are typically in the 'deep MOND' regime, and only transition to high accelerations fleetingly at very close distances. The results of a MOND based analysis are masses which are approximately 1% of their usual Newtonian estimates. This

Figure 5.6: Mass estimate contours for the LG using MOND with the simple and standard interpolation functions. The masses are significantly lower than the bayonic masses of MW and M31, by approximately an order of magnitude. This is consistent with [66] who find that for MOND to be consistent with the TA, the galaxies need to be on their second pass.



presents significant tension with observational results, which suggest that the usual baryonic mass of the MW and M31 significantly exceed this (making up approximately 10% of the Newtonian estimate). In order to resolve this issue, the MW and M31 need to be on their second pass, with a fly-by event in the past as in [66].

It should be noted that MOND is a non-linear theory which depends on the *absolute* magnitude of the acceleration, and hence accelerations from internal and external fields cannot be added independently. Since the matter inside galaxies is necessarily rotating, and towards the centre of the galaxy we are in a quasi-Newtonian regime whilst the outside is in the 'deep-MOND', the interior and exterior extremes of the galaxy will couple differently to the external field. As such, one would generically expect a structure dependent coupling to macroscopic fluid objects (such as galaxies or gas clouds) based on the acceleration profile of the objects. This may change the effects of MOND quite significantly.

5.8 Joint Analysis with Cosmological Probes

We can marginalise over the cosmological parameters relevant to the TA by using the results of sampling analyses on cosmological datasets such as LSS or the CMB. The key parameters which are important are $\Omega_{\rm DE}$, $\Omega_{\rm m}$, h, and w_{DE}, which together provide the age of the universe t_u and the necessary modifications to the gravitational equation.

The marginalised local group mass is

$$\langle M_{\rm LG}(r,v_r)\rangle = \int M_{\rm LG}(r,v_r|\Omega_m,\Omega_{\rm DE},w_{\rm DE},h)P(\Omega_m,\Omega_{\rm DE},w_{\rm DE},h)\,d^4\vec{C}, \quad (5.72)$$

where $\vec{C}$ is a vector representing the cosmological parameters $\{\Omega_m,\Omega_{\rm DE},w_{\rm DE},h\}$. (Note that we need to normalise the posterior distribution from the analysis.) We look at the probability distribution of the $M_{\rm LG}$ based on the probability distributions of the dynamical and cosmological variables by treating it as a derived parameter. The dependence on cosmology may be washed out by the observational uncertainties for the local group dynamics so we shall consider the cosmological uncertainties in isolation. In the absence of data, we take a theoretical posterior distribution, obtained by the procedures described in Part I using a slightly simplified case: the Euclid photometry (with fixed redshift distributions) and CMB TT information and a fiducial cosmology of $\Omega_{fld}=0.7$, $h=0.7$, and $w=-1$. The likelihood (equation 3.26) is sampled using the nested sampler PLINY to calculate the posterior distribution given uniform priors over the cosmological parameters. This yields the cosmological posterior shown in Fig 5.7, where M_{LG} has been included as a derived parameter. We can clearly see the dependencies of M_LG on the cosmological parameters by the slopes of the ellipses. By calculating the LG mass for the different values of Ω_{fld}, h, and w, and using the posterior distribution shown in Fig 5.7 to assign weights, we may calculate the probability distribution over the mass of the LG shown in Fig 5.8. The marginalised mass estimate is $M_{LG}=5.00\times10^{12}M_{\odot}$.

It's clear that the uncertainty in M_{LG} derived from the bounds on cosmological parameters in wCDM is smaller by an order of magnitude compared to those deriving from the observed dynamics of the galaxies or the TA_Λ itself (see tables 4.4 and 4.5 for the results of the previous chapter and a survey of other results in the literature). Hence, assuming no non-GR physics, the impact of unknown cosmology is minimal on the mass estimate. In principle, scalar field theories can have a much greater impact (such as the symmetrons in this chapter), although cosmological analyses rarely select specific models at the level of the action, so consistency checking is difficult without highly specialised calculations for individual models.

Figure 5.7: Probability contours for cosmological parameters obtained using the Euclid like photometry setup for galaxy number counts and CMB-TT information. The important parameters are Ω_{fld}, h, and w, which affect our LG mass estimate.

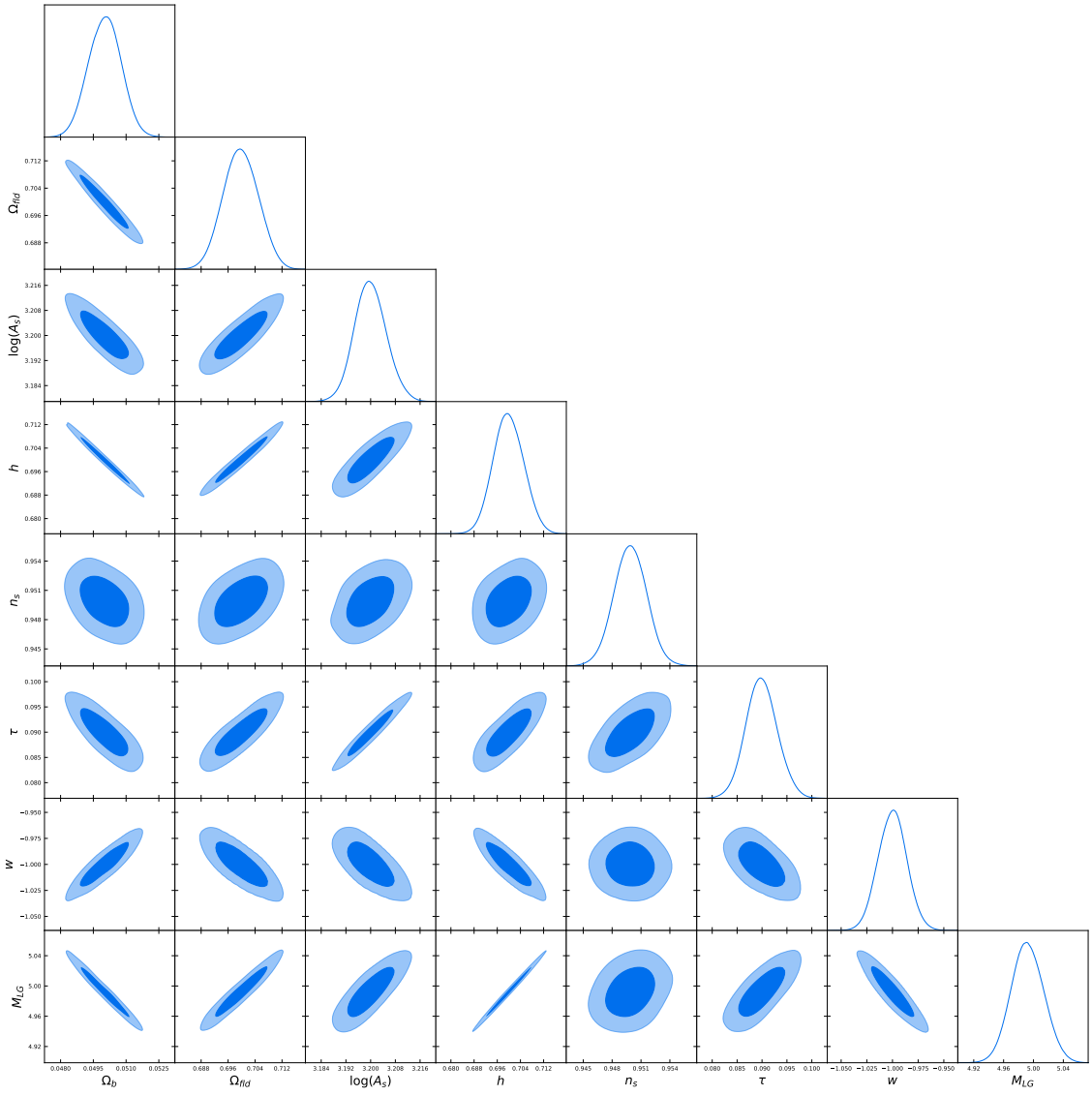
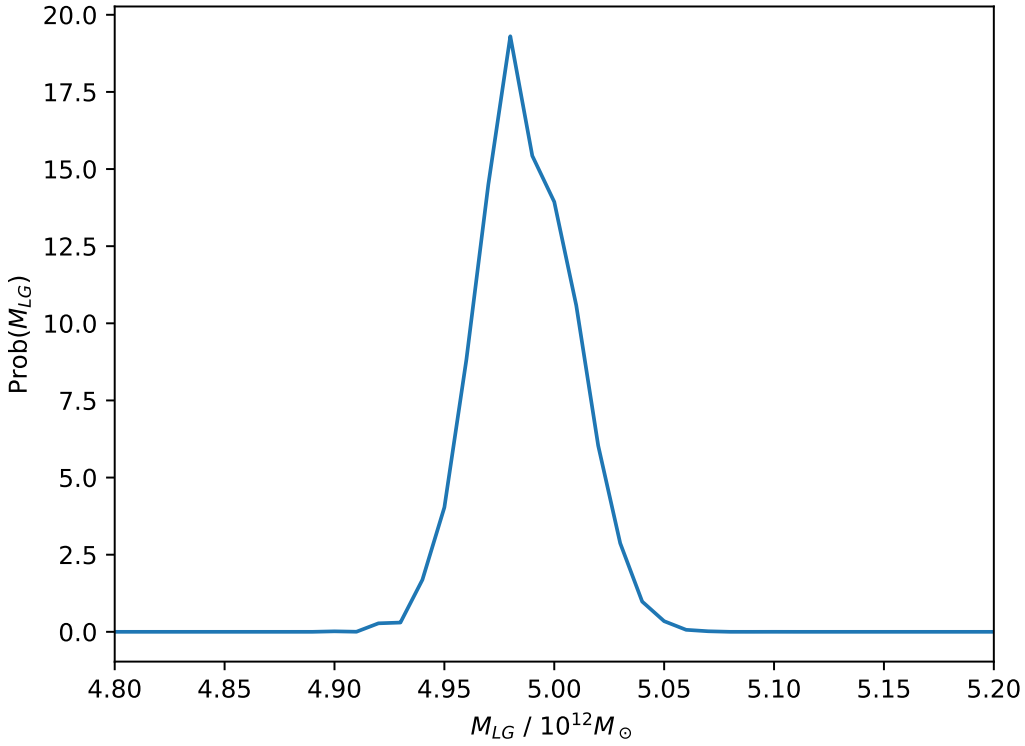


Figure 5.8: Probability distribution over M_{LG} given the posterior in Fig 5.7.



It is worth bearing in mind that such tight constraints can be misleading. Tensions between estimates for H_0 (summarised in [75]) would lead one to consider much wider probability distributions, which could easily change the value of the LG mass by $2 \times 10^{11} M_\odot$ or more. The fiducial cosmological model used in this section has $h = 0.7$, as opposed to the Planck like value of 0.67. As a result, the mass estimate is boosted from $4.8 \times 10^{12} M_\odot$ to $5.0 \times 10^{12} M_\odot$. Whilst this is still below the level of other uncertainties in the TA, it is considerably higher than the tight posterior distributions on the cosmological parameters from this forecast would suggest.

5.9 The Local Group as a Laboratory for Dark Energy and Modified Gravity

In principle, the LG and other similar Mpc scale systems could be used to make inferences about gravity and dark energy, by reversing the approach we have so far taken. If, by some independent means (such as strong gravitational lensing or stellar streams), one *knew* the mass of the system, one could then constrain gravity by looking at the mass estimates of the Local Group. We do this to some extent when we look at MOND or symmetrons. In this case, the mass of the galaxies is understood reasonably well by estimating the baryonic mass of the Milky Way and Andromeda; in order to be viable, one must find a MOND or symmetron theory which provides a consistent mass estimate. Stellar streams may also be used to estimate the mass of the Milky Way through gravitational effects, but on a much smaller scale. The consistency between these scales would be the key to understanding the gravitational theory. We can see from the contour plots in the previous sections that each of these theories has free parameters which are degenerate with respect to the LG mass estimate; in order to constrain the theory more fully one would have to break this degeneracy with other observations.

5.10 Discussion

We summarise some examples of LG mass estimates in table 5.2 for a variety of models and parameters. We find that both of our dark matter free theories – MOND and symmetrons – produce mass estimates which are inconsistent with estimates of the baryonic mass of the Milky Way and Andromeda. Although, in the case of the symmetrons, the modelling is over simplified, it places significant pressure on these theories as replacements for dark matter which are consistent on scales larger than the galaxies themselves. Different models of dark energy in unmodified GR produce relatively weak effects on the LG mass estimate, although values of h at the extremes of the observed range produce changes in M_{LG} which are of the same order of magnitude as uncertainties in observables such as radial separation, velocity, and the uncertainty due to the considerable simplifications in the model

Table 5.2: Examples of LG mass estimates for different models and parameters. Parameters in the symmetron model are given here in Planck units, and are based on [65]. The variation between models can be quite large, even when remaining in ΛCDM due to the tension over h.

Model	Parameters	M_{LG} / $10^{12} M_\odot$
ΛCDM	$h = 0.67$, $\Omega_\Lambda = 0.7$	4.80
ΛCDM	$h = 0.632$, $\Omega_\Lambda = 0.7$	4.55
ΛCDM	$h = 0.764$, $\Omega_\Lambda = 0.7$	5.43
wCDM	$h = 0.67$, $\Omega_f = 0.7$, $w = -1.1$	4.79
wCDM	$h = 0.67$, $\Omega_f = 0.7$, $w = -0.9$	4.81
Symmetron + Λ	$\mu = 2.46 \times 10^{-58}$ GeV, $\lambda = 10^{-109}$, $S = 0.015$	3.34
MOND	$a_0 = 1.2 \times 10^{-10}\,\mathrm{m\,s^{-2}}$, $t_u = 13.69$ Gyr	0.027

itself. This means that we cannot ignore cosmology when calculating the LG mass, and we should always consider these estimates in their proper context. Whilst the effects on the LG mass are not extreme enough to produce masses which we can claim are inconsistent with other observations on the LG – and thus we cannot turn this analysis around and place a limit on ΛCDM – it appears that cosmology could be an important component of precise estimates of the LG mass in future.

Chapter 6

General Conclusions

6.1 Summary and General Conclusions

Here we aim to draw together the various threads of this thesis, and demonstrate the connections between the different areas of physics that we have considered.

6.1.1 Constraining redshift with cross-correlations

We began by considering cosmological statistics in ΛCDM. This provides the most important benchmark for new methods in cosmological inference. We showed that the cross-correlation functions between redshift bins contain not only information about the cosmology, but information about the overlap of the redshift distributions. We found that using just 3 spectroscopic bins across the redshift range, we could constrain the mean of a gaussian bin to 3×10^{-3}, and its variance to the level of 10^{-3}. Such tight constraints on redshift prevent the redshift errors from propagating significantly into the cosmological parameters, and thus the degeneracies between redshifts and cosmology vanish. Just as crucially, we demonstrated that systematically incorrect redshift distribution can bias the cosmological parameters, so this method improves both precision and accuracy.

We next continue in the same vein, to test the method in a more realistic survey scenario as well as to forecast constraints on modified gravity from cross-correlations. Using a denser and broader spectroscopic sample, we show that redshifts that overlap with a number of redshift distributions are extremely well constrained, with $\Delta\mu = 10^{-3}$ and $\Delta\sigma$ reaching as low as 5×10^{-4}. This meets the

requirements set in [45], for example, for doing precision cosmology in the future. Despite introducing a large number of a parameters to one's sampling volume, the strong constraints on all of the parameter lead to reasonably good behaviour.

6.1.2 Cosmological inference with galaxy and CMB datasets

We found that we obtain strong cosmological constraints from a cross-correlation analysis. By simultaneously constraining redshifts with spectroscopic samples, we not only evade bias but recover posterior contours very extremely close to those obtained with a fixed redshift distribution. This means that we can address the uncertainties in photometric redshifts in a consistent way, without degrading our cosmological results when the redshift parameters are marginalised over. The constraints on cosmological parameters when going from our small DES sample to the full Euclid sample is very promising, and the cosmological parameters are recovered with uncertainties at the percent level or better. When w is included, however, the degeneracies between w and other cosmological parameters are very significant and cause our constraints on the standard cosmological parameters to loosen considerably (although the use of modified gravity may also contribute to this). A potentially valuable dataset to combine into our analysis in such a case would be supernovae, to rigidly constrain the background evolution at a level better than we are able to with our current limited set of observables.

6.1.3 Generalised approaches to modifying gravity

The journey from ΛCDM to modified gravity is a central theme to this work as a whole, and the consistency between physical theories on different scales is hotly debated. GR is enormously successful in cosmology, astrophysics, and even terrestrial physics, but there is room in the cosmological probes for doubt. We show that LSS and weak lensing, as observed by Euclid, will be a powerful tool for constraining dark energy. On the other hand, we also saw the sensitivity of the results to the exclusions of α_T, and therefore are forced to confront the fact that our modified gravity predictions are not necessarily robust or generic, but rather that the observable effect depend heavily on the phenomenological choices which are made. This

makes comparison with other, earlier tests of modified gravity particularly challenging, as we see that the available parameter space can change radically due to stability criteria. We explored the stability of the parameter space in the absence of α_T, and find our modified models are heavily restricted by its absence.

A general approach to modified gravity is also taken in chapter 5, by calculating generic spherical solutions for weakly perturbed scalar fields around galaxy size haloes. We can therefore investigate how the Local Group mass estimate changes depending on the properties of the scalar field solution, without having to explore individual theories and solve for different couplings and potentials. The range of possible behaviour is very broad, and can be used to inspect what requirements are demanded of our scalar field solution for consistency with other measurements.

6.1.4 Connecting the Local Group to different scales in the universe

We have explored novel aspects of a relatively straightforward problem: an argument to estimate the Local Group mass as the sum of the masses of MW and M31 (and any bound satellites which move with them). Through the use of ANN and numerical simulations, we have demonstrated evidence that the mass estimate can be improved by including information about the local velocity shear, connecting the Local Group dynamics to larger external flows on the scale of a several Mpc. We furthermore consider the Local Group mass as a derived parameter of the cosmology, utilising the kind of posterior distributions developed in the first part of the thesis. This brings the Local Group into contact with the very largest scales, and we can see that the uncertainty in h in particular casts a shadow of doubt over mass estimate of the same order of magnitude as other observational inputs such as the radial separation or the velocities. Understanding the Local Group dynamics brings together the background evolution of the universe, the nature of dark energy, and Newtonian dynamics into a simple problem which shows many facets. One of the great goals in physics nowadays is to connect scales together. We identify the Local Group as a source of tension for dark matter free theories which seek to stabilise rotation curves such as MOND or attractive scalar fields. The far field form of the

force modifications cannot mimic the additional dark matter mass out to large distances, and the mass estimates come out too high (symmetrons) or too low (MOND) to compensate.

6.2 Future Work

6.2.1 Physics on Mpc scales

The Local Group makes a nice playground for Mpc scale physics, and one that is close to home. Unfortunately we have only one, and its properties are not always well understood! Further study of similar effects in galaxy clusters and other similar scaled systems may demonstrate some of the same sensitivities as our local analyses, and thus lend some statistical weight. I believe it will be increasingly important in the future to connect modified gravity theories on scales of galaxies, clusters, the cosmic web, and cosmology in order to subject it to rigorous consistency tests. Many theories are tailored towards producing behaviour on a particular scale (e.g. cosmic acceleration, or stabilising galaxy rotation curves), and thus a multi-scale analysis of theories is necessary to truly understand their impact beyond the task that they have been created to do. Even without scalar fields, Mpc scales contain interesting aspects. Understanding the Local Group offers many exciting opportunities for exploring the behaviour and formation of galaxies and galaxy clusters. Constrained simulations are being used to reconstruct the history of our local universe [109]. It is possible that an ensemble of realisations could be used with statistical methods to better understand the properties of the galaxies and flows around us [108][81], especially when combined with more advanced algorithms.

6.2.2 Preparing modified gravity for the future

Observing modified gravity (or its absence) is a major task for the next generation of experiments. In order to explore this well, we need to be confident that our models are covering the available parameter space reasonably well. To avoid having contours biased by pathological models with gradient instabilities (which might have a perfectly reasonable level of effect on the cosmological observables), we ought to use a more robust parameterisation. The inclusion of a linear parameterisation

such as $\alpha_i = c_i \frac{\Omega_\Lambda}{\Omega_\Lambda^0} + b_i$ could help to keep the background stable by avoiding the part of the history where c_s^2 crosses zero. This however suffers from a lack of physical motivation, and is not particularly likely to closely represent a physical theory derived from a reasonable action. Parameterising the evolution of the α-functions directly, by expanding them as polynomials or fitting an interpolated spline with a small number of points, would lend us considerable flexibility but at the cost of an enormous and difficult to constrain parameter space. It would be interesting what constraints we could get on the three remaining α-functions using the full weight of the next generation of galaxy surveys, supernovae, and CMB measurements.

From a theoretical point of view, now that GW170817 has removed α_T from our consideration, we are now faced with a greatly simplified scalar-tensor action. It would be useful to see if the class of functions available to use has shrunk enough that some reasonable ansatz may be made which is motivated by, and close to, the evolution of genuine scalar field theories. This would remove the need to constrain the time evolution directly, but would certainly force us to look at subclasses of the general scalar tensor action (those which share a similar time dependence). The advantage of fully understanding the action which you are constraining however would be significant in interpreting any results we might retrieve from future experiments.

A particularly useful forecast for modified gravity analysis would be to explore the capabilities of different modified gravity parameterisations for detecting the effects of given full theories. Detailed mock data should be created using known modified gravity models, and taken as the data for a pipeline which performs a general modified gravity analysis. The ability for parameterisations to mimic a wide variety of theories is crucial to the future of observational modified gravity, and needs to be quantified.

6.2.3 Multi-probe analysis with the next generation of experiments

Analyses featuring multiple probes are now the norm in cosmology, and cross-correlations between datasets are becoming increasingly important in order to

achieve the precision that we desire. Weak lensing in particular is a focus of much interest, given the unprecedented size of the data sets which will come from Euclid and LSST, giving us the opportunity to beat back the large uncertainties associated with weak lensing by statistical power. At UCL, we are already working considerably on cross-correlations projects on BOSS data, and working towards building pipelines for more intricately combined datasets. The future will require laying the foundation so that we can combine the probes we want, cross-correlate those which will be beneficial, and jointly constrain our photometric redshifts seamlessly and efficiently. In order to be ready for the next generation, the kind of analyses performed in this paper must be moved into the realm of real data, and potentially include non-linear modelling of the $C(l)$s out to higher l to make the most of the data which is available to us. The methods which are employed to exploit the data available to us – from BOSS, Planck, or DES – can also inform continuing forecasts as we hone our methods. Whilst we have shown that modified gravity is susceptible to constraint from using number counts, weak lensing shear (both $l \leq 250$), and CMB-TT ($l \leq 3000$), this is really only the beginning. There is a wealth of information coming, ready to be exploited.

Bibliography

[1] Riess A., et al., 1998, Astron. J. 116, 1009-1038

[2] Perlmutter S. et al., 1999, ApJ 517, 565

[3] Perlmutter S. et al., 1998, Nature 391, 51

[4] Weinberg S., 1989, Rev. Mod. Phys. 61, 1

[5] Peebles P.J.E & Ratra Bharat, 2003, Rev. Mod. Phys. 75, 559

[6] Martin J., 2012, Comptes Rendus Physique 13, 6, 566

[7] Zaroubi S., 2012, 'The Epoch of Reionisation' in 'The First Galaxies', ASSL 396

[8] Atek H. et al., 2015, arXiv:1509.06764

[9] Guth A.H., 1981, Phys. Rev. D 23, 347

[10] Lesgourgues J., Pastor S., 2006, Physics Reports 429, 307

[11] Lesgourgues J., Pastor S., 2012, Adv. High Energy Phys., 608515

[12] Thomas S.A., Abdalla F.B., Lahav O., 2010, Phys. Rev. Lett. 105, 031301

[13] Giusarma E., Gerbino M., Mena O., Vagnozzi S., Ho S., Freese K., 2016, Phys. Rev. D 94, 083522

[14] Markevitch M. et al., 2004, ApJ 606, 819

[15] Bassett B.A., Hlozek R., 2009, arXiv:0910.5224

[16] Ade P.A.R. et al., 2014, Phys. Rev. Lett., 112, 241101

[17] Gott J.R., Colley W.N., 2017, arXiv:1707.06755

[18] Grayson J.A. et al., 2016, arXiv:1607.04668

[19] Raby S., 2006, arXiv:hep-ph/0608183

[20] Bojowald M., 2011, Nucl. Phys. A 862, 98

[21] Rovelli C., Vidotto F., 'Covariant Loop Quantum Gravity', 2015, Cambridge University Press

[22] Bartolo N., Komatsu E., Matarrese S., Riotta A., 2004, Physics Reports 402, 103

[23] Yoo J., Watanabe Y., 2012, Int. J. Mod. Phys. D21, 1230002

[24] Knobel C., 2012, arXiv:1208.5931

[25] Peacock J.A., 'Cosmological Physics', 1999, Cambridge University Press

[26] Hobson M., Efstathiou G., & Lasenby A., 'General Relativity', 2006, Cambridge University Press

[27] Vitagliano V., Sotiriou T.P., Liberati S., 2011, Annals Phys. 326, 1259

[28] Sotirious T.P., Liberati S., 2007, Annals Phys. 322, 935

[29] Peter P. & Uzan J., 'Primordial Cosmology', 2009, Oxford University Press

[30] Peebles P. J. E., 'The Large-Scale Structure of the Universe, 1980, Princeton University Press

[31] Kilbinger M., 2015, Rep. Prog. Phys. 78, 086901

[32] Troxel M.A., Ishak M., 2015, Physics Reports 558, 1

[33] Joachimi B., Mandelbaum R., Abdalla F.B., Bridle S.L., 2011, Astronomy and Astrophysics 527, A26

[34] Meneghetti M., Lecture Scripts, 'Introduction to Gravitational Lensing'

[35] Abdalla F., Rawlings S., 2008, MNRAS, 381, 1313

[36] Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2008, MNRAS, 417, 1891

[37] Banerji M., Abdalla F., Lahav O., Lin H., 2008, MNRAS, 386, 1219

[38] Benjamin J. et al, 2010, MNRAS, 408, 1168

[39] Blake C., Ferreira P. G., Borrill J., 2004, MNRAS, 351, 923

[40] Blake C. Collister A., Bridle S., Lahav O., 2007, MNRAS, 374, 1527

[41] Blas D., Lesgourgues J., Tram T., 2011, JCAP, 07, 034

[42] Bucher M., Moodley K., Turok N., 2002, Phys. Rev. D 66, 023528

[43] Clerkin L. et al., 2014, MNRAS 448, 1389

[44] Huterer D., 2001, ApJ, 555, 547

[45] Huterer D., et al., 2004, ApJ, 615, 595

[46] Kirk D., Lahav O. et al., 2013, MNRAS, 438, 2218

[47] Hikage C., Yamamoto K., 2016, MNRAS, 255, L77

[48] Kaiser N., 1987, MNRAS, 227, 1

[49] Padmanabhan N. et al., 2007, MNRAS, 378, 852

[50] Lahav O., Kiakotou A., Abdalla F. B., Blake C., 2009, MNRAS, 405, 168

[51] Matthews D., Newman J., 2010, ApJ, 721, 456

[52] McQuinn M., White M., 2013, MNRAS, 433, 2857

[53] Menard B. et al., 2011, arXiv:1303.4722v2

[54] Newman J., 2008, arXiv:0805.1409

[55] Newman J. A. et al., 2012, Astroparticle Physics, 63, 81

[56] Peebles P. J. E., 1973, ApJ, 185, 413

[57] Rhodes J. et al., 2014, arXiv:1309.5388v3

[58] Sadeh I., Abdalla F., Lahav O., 2015, arXiv:1507.00490v1

[59] Schmidt S. et al., 2013, MNRAS, 431, 3307

[60] Schulz A.E., 2006, ApJ, 724, 1305

[61] SDSS Collaboration, 2008, Astron. J, 142, 72

[62] Thomas S. A., Abdalla F., Lahav O., 2012, MNRAS, 412, 1169

[63] Zheng H. & Zhang Y., 2012, Proc. SPIE 8451, Software and Cyberinfrastructure for Astronomy II, 845134

[64] Burrage C., Copeland E. J., Hinds E. A., 2015, JCAP 03 042

[65] Burrage C., Copeland E.J., Milington P., 2017, Phys. Rev. D, 95, 064050

[66] Zhao H., et al., 2013, A&A, 557

[67] Brax P., 2012, arXiv:1211.5237v1

[68] Mota D. F., Shaw D. J., 2007, Phys. Rev.D, 75, 063501

[69] Milgrom M., 1983, ApJ, 270, 365

[70] Scarpa R., 2006, AIP Conference Proceedings, 822, 253

[71] Lidsey J.E., Liddle A.R., Kolb E.W., Copeland E.J., Barreiro T., Abney M., 1997, Rev. Mod. Phys. 69, 373

[72] Sahlén M., Liddle A. R., Parkinson D., 2005, Phys. Rev. D, 72, 083511

[73] Tsujikawa S., 2013, Class. Quant. Grav., 30, 214003

[74] Ratra B., Peebles P. J. E., 1988, Phys. Rev. D, 37, 3406

[75] Jackson N., 2015, Living Reviews in Relativity, 18:2

[76] Planck Collab, 2016, A&A 594, A13

[77] Blas D., Lesgourgues J., Tram T., 2011, JCAP, 07, 034

[78] Freedman W. L. et al., 2012, ApJ, 758, 24

[79] Braatz J. et al, 2012, Proc. IAU, 289, 255

[80] Binney J., Tremaine S., 2008, Galactic Dynamics, Princeton University Press

[81] Carlesi E., Hoffman Y., Sorce J., Gottlöber S., 2016, MNRAS 465, 4886

[82] Collister A., Lahav O., 2004, Publ. Astron. Soc. Pac., 116, 345

[83] Courteau S. et al., 2014, Rev. Mod. Phys, 86, 47

[84] Cybenko G., 1989, Math. Control Signals Systems, 2, 303

[85] de Vaucouleurs G., 1975, Galaxies and the Universe, Chapter 14, Nearby Groups of Galaxies, University of Chicago Press

[86] Diaz J.D., Koposov S.E., Irwin M., Belokurov V., Evans M.W., 2014, MNRAS 443, 1688

[87] Einasto J., Lynden-Bell D., 1982, MNRAS, 199, 67

[88] Forero-Romero J. E., González R., 2015, ApJ, 799, 45

[89] Forero-Romero J. E., Contreras S., Padilla N., 2014, MNRAS, 443, 1090

[90] González R.E., Kravtsov A.V., Gnedin N.Y., 2014, arXiv:1312.2587v2

[91] Hoffman et al., 2012, MNRAS, 425, 2049

[92] Kahn F.D. and Woltjer L., 1959, ApJ, 130, 705

[93] Knebe A. et al., 2011, MNRAS, 415, 2293

[94] Klypin A. et al., 2014, arXiv:1411.4001v2

[95] Kroeker T.L., Carlberg R.G., 1991, ApJ, 376, 1

[96] Lahav O., Lilje P.B., Primack J.R., Rees M.J., 1991, MNRAS, 251, 128

[97] Lahav O., Naim A., Sodré L., Storrie-Lombardi M. C., 1996, MNRAS, 283, 207

[98] Li Y-S., White S., 2008, MNRAS, 384, 1459

[99] Visser M., 2005, Gen. Rel. Grav. 37, 1541

[100] Libeskind N., Hoffman Y. et al., 2012, MNRAS, 428, 2489

[101] Libeskind N., Hoffman Y. et al, 2015, MNRAS, 452, 1052

[102] Libeskind N. et al., 2013, MNRAS, 428, 2489

[103] Coutinho B.C. et al., 2016, arXiv:1604.03236

[104] Fisher J.D., Faltenbacher A., Johnson M.S.T., MNRAS, 458, 1517

[105] Forero-Romero J.E., Hoffman Y., Gottloeber S., Klypin A., Yepes G., 2009, MNRAS, 396, 1815

[106] Hoffman Y. et al., 2012, MNRAS, 425, 2049

[107] Tully B.R., Courtois H., Hoffman Y., Pomarède D., 2014, Nature 513, 71

[108] Libeskind N., Yepes G., Knebe A., Gottloeber S., Hoffman Y., Knollman S.R., 2010, MNRAS, 401, 1889

[109] Carlesi E. et al., 2016, MNRAS 458, 900

[110] van den Bergh S., 2003, arXiv:astro-ph/0305042

[111] DES Collab., 2015, ApJ 813, 109

[112] Lynden-Bell D., 1981, The Observatory, 101, 111

[113] MacKay D., 2003, Information Theory, Inference, and Learning Algorithms, C.U.P.

[114] Metuki O., Libeskind N., Hoffman Y., Crain R.A., Theuns T., 2015, MNRAS 446, 1458

[115] Nuza S.E., Kitaura F-S., Heß S., Libeskind N., Müller V., 2014, MNRAS, 445, 988

[116] Pahwa I. et al., MNRAS, 457, 695

[117] Partridge C., Lahav O., Hoffman Y., 2012, MNRAS, 436, L45

[118] Peñarrubia J., Ma Y-Z., et al., 2014, MNRAS, 443, 2204

[119] Peñarubbia J., Gómez F.A. et al., 2015, MNRAS, 456, L54

[120] Phelps S., Nusser A., Desjacques V., 2013, ApJ, 775, 102

[121] Planck Collaboration, 2015, arXiv:1502.01589v2

[122] Raychaudhury S., Lynden-Bell D., 1989, MNRAS, 240, 195

[123] Riebe et al, 2013, AN, 334, 691

[124] Sadeh I., Abdalla F., Lahav O., 2015, arXiv:1507.00490v1

[125] Salomon J.B., Ibata R.A. et al., MNRAS, 456, 4432

[126] van der Marel R. et al., 2012, ApJ, 753, 8

[127] van der Marel R., Guhathakurta P., 2008, arXiv:0709.3747v2

[128] Yepes G., Gottlöber S., Hoffman Y., 2013, arXiv:1312.0105v1

[129] Brans C., Dicke R.H., 1961, Phys. Rev. 124, 925

[130] Avilez A., Skordis C., 2014, Phys. Rev. Lett. 113, 011101

[131] Clifton T., Ferreira P.G., Padilla A. Skordis C., 2012, Physics Reports 513, 1, 1

[132] Zumalacárregui M., Bellini E., Sawicki I., Lesgourges J., Ferreria P., 2016, JCAP 1708, 08, 19

[133] Horndeski G.W., 1973, Internation Journal of Theoretical Physics, Vol. 10, No. 6, 363

[134] Deffayet C., Steer D.A., 2013, Classical and Quantum Gravity, 30, 21

[135] Renk J., Zumalacárregui M., Montanari F., Barreira A., 2017, JCAP 10, 020

[136] Renk J., Zumalacárregui M., Montanari F., 2016, JCAP 07, 040

[137] Bellini E., Cuesta A.J., Jimenez., Verde L., 2016, JCAP 06

[138] Alonso D., Bellini E., Ferreira P.G., Zumalacárregui M., 2016, Phys. Rev. D95, 6, 063502

[139] The Dark Energy Survey, 2018, "https://www.darkenergysurvey.org/collaboration-and-sponsors", accessed 27/03/18

[140] DES collab., 2005, arXiv:astro-ph/0510346

[141] Simpson F. et al., 2013, MNRAS 429, 3, 2249

[142] Laureijs R. et al., 2011, ESA/SRE(2011)12, arXiv:1110.3193v1

[143] Amendola L. et al., 2016, arXiv:1606.00180

[144] Sartoris B. et al., 2015, arXiv:1505.02165

[145] LSST Dark Energy Science Collaboration, 2012, arXiv:1211.0310 [astro-ph.CO]

[146] Hojjati A., Pogosian L. Silverstri A., Zhao G.B., 2014, Phys. Rev. D 89, 083505

[147] Banerjee A., Jain B., Dalal N., Shelton J., 2017, arXiv:1612.07126

[148] Baker T., Bellini E., Ferreira P.G., Lagos M., Noller J., Sawicki I., 2017, Phys. Rev. Lett. 119, 251301

[149] Ezquiaga J.M., Zumalacárregui M., 2017, Phys. Rev. Lett. 119, 251304

[150] Abbott B.P. et al., 2017, Phys. Rev. Lett. 119, 161101

[151] LIGO Scientific Collab., Virgo Collab., Fermi Gamma-Ray Burst Monitor, INTEGRAL, 2017, ApJ Letters, 848:L13

[152] Joachimi B., Bridle S.L., 2010, Astronomy and Astrophysics 523, A1

[153] Bellini E., Sawicki I., 2014, JCAP 07, 050

[154] Will C.M., 1993, 'Theory and Experiment in Gravitational Physics', Cambridge University Press

[155] Li E., 2016, arXiv:1612.07781

[156] DES collab., 2017, FERMILAB-PUB-17-294-PPD, arXiv:1708.01530

[157] DES collab., 2017, FERMILAB-PUB-17-279-PPD, arXiv:1708.01538

[158] Fry J.N., 1996, ApJ 461, L65

[159] Matarrese S., Coles P., Lucchin F., Moscardini L., 1997, MNRAS 286, 115

[160] Tinker J.L. et al., 2010, ApJ 724, 878

[161] Skilling J., 2006, Bayesian Analysis, No. 4, 833

[162] Feroz F., Hobson M.P., Bridges M., 2009, MNRAS 398, 1601

[163] Handley W.J., Hobson M.P., Lasenby A.N., 2015, MNRAS 450, L61

[164] Mukherjee P., Parkinson D., Liddle A.R., 2006, ApJ, 638, L51

[165] MacKay D.J.C., 'Information Theory, Inference, and Learning Algorithms', 2003, Cambridge University Press

[166] Lahav O., Naim A., Sodré L., Storrie-Lombardi M.C., 1996, MNRAS 283, 207